

统计生态学

——方法和计算入门

Jhon A. 拉德维格

James F. 蓝诺兹

内蒙古大学出版社

58.18
327

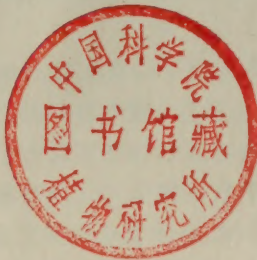
统计生态学

——方法和计算入门

[澳] Jhon A. 拉德维格[✓] 著
[美] James F. 蓝诺兹

李育中 王 炜 裴 浩 译
刘钟龄 校

李育中
↓
李育中



内蒙古大学出版社
1990

中科院植物所图书馆



S0015494

267.1



译者的话

我们这本初稿译物的目的是为生态学中现代统计论题的一些介绍。所

近年来,随着计算机的推广与普及,统计和多元分析的采用,促进了数量生态学的迅速发展,国内也出版了几本译著,如《植物群落排序》、《植物群落分类》、《群落生态学的多元分析》。但比较系统地、详细地讲解统计生态学的方法和计算过程的书还没有,因而很难进行这方面的工作。基于这一点,我们翻译了《统计生态学——方法和计算入门》。该书全面地比较了各种统计生态学的方法,并在每个论题里选出最佳几种方法,逐步地详细介绍,几个简单例子贯通全书,并配有一套软盘,读者阅读完各种方法后可以用书中的例子或自己的数据直接上机操作,这样可以加深对各种统计生态学方法的透彻理解。该书叙述简洁,步骤清晰,易于阅读,便于计算。

本书共分24章,1—7章、17—21章、24章由李育中译,8—12章由王炜译,13—16章、22章、23章由裴浩译。刘钟龄教授校订。

限于水平,翻译中一定存在不少缺点和错误,欢迎读者批评指正。

统计生态学由两部分组成,首先,我们给出了在群落生态学中关于数据收集的基本原理的一些例子,简单地显示了如何取样和怎样将数据组织成矩阵的形式。当然,统计生态学的一个重要组成部分,这一学科有很多可利用的取样方法。在第一部分里我们的目的就是强调它的重要性。

这本书的其它部分包括:空间格局分析,种多度关系,种重要模型,群落分类和排序。最后的一部分我们称为群落的关系。在每部分开头的一章里我们以背景信息开始为学生提供这一学科的一些课程材料。基于这一目的,我们列了一个所选择的文献表。这样使读者能够学习到在生态学研究某些方法使用的一些特定例子。我们使这个部分的学生阅读和评论许多这样的文章,并且相信让学生们接触这些文献是学习过程中十分重要的一个部分。

在每部分所包含的特定方法中剩余几章里按以下形式安排。我们给出必要的预备信息,其中包括一些参考文献,接下来包括的是计算步骤。这些步骤是以一步一步的翻译者解释的。我们使用这种渐进式的方法是由于后面附着的计算实例也是以同样的方式。1990. 12 这样就使得后面的计算能很快地参照前面清晰的计算步骤和方程。由于我们的目的是帮助学生们理解所包含的计算过程,所以这些计算实例也非常简单(利用人为的数据组)。在某些情况下,这种折衷办法既简单又带着生态-统计的真实性。在这本教科书里,我们尽量点出,哪里存在着这种重要性。在这些示例的后边,我们进一步提供了使用计算机计算的一些其它例子。

在每一章方法的后面我们还提供了一节附加讨论和推荐的总结。附加的论题由简单的陈述和选择的文献组成,其目的是强调一下在前面的介绍中没有提到的一些方法。同学们会发现这些论题是不完备的,更确切地说,我们的目的是提出一些问题和论题,这些也许能引起对每种方法有基本了解的同学的兴趣。这些参考文献也为进一步学习提供了指南。

前言

我们这本初级读物的目的是为初学者提供在群落生态学中现代统计论题的一些介绍。所提供的各种方法,其广度和深度是以我们在教没有统计生态学基础和仅有一点统计基础的学生为基点的。对于初学的学生一个学期的课程,我们不想对于所有可利用的方法作彻底的叙述。相反的,我们是想引导学生选择论题的范围,其中的一些是:(1)在生态学历史上是重要的(如极点排序);(2)生态学家公认的(如多样性指数);(3)分析生态格局强有力的工具(如多维调节)。我们所提供的方法可能分属于这三种类型但不一定是必须属于这三种。例如,我们提供的多样性指数主要是由于它的广泛普及性;在生态学的文献中同学们将会经常见到多样性指数出现,所以我们试图讲解一下它们正确地 and 错误地使用情况。在深度方面,本书试图避免过分的简单化和过分数学化。

从历史的角度来看,生态学在很大程度上是由描述性的学科演变为现在这种较高定量化领域的状态,因此,这就要求研究人员不断地学习新的而且经常是复杂的、数量化的技术。尽管在最近出版的关于数量生态学的一些书籍(如 Legendre 和 Legendre 1983, Orloci 和 Kenkel 1985, Pielou 1984)提供了有关这方面理论和一些特定统计方法应用的细节,但我们仍感到有必要为初学者提供这一领域广度适当的方法,这就是我们这本书的目的所在。

统计生态学这本书由 7 个部分组成,首先,我们给出了在群落生态学中关于数据收集基本原理的一些总的看法和简单地展示了如何取样和怎样将数据组织成矩阵的形式。当然取样是统计生态学的一个极其重要的部分,这一学科有很多可利用的取样方法。在第一部分里我们的目的就是重新强调它的重要性。

这本书的其它部分包括:空间格局分析,种多度关系,种重叠模型,群落分类和排序,最后的一部分我们称为群落的解释。在每部分开头的一章里我们以其背景信息开始为学生提供这一学科的一些透视材料。基于这一目的,我们列了一个所选择的文献表,这样使得感兴趣的学生们能够学习在生态学研究某些方法使用的一些特定例子。我们使上这个课的学生阅读和评论许多这样的文章,并且相信让学生们接触这些文献是学习过程中十分重要的一个部分。

在每部分所包括的特定方法中剩余几章里按以下形式安排。我们给出必要的预备信息,其中包括一些参考文献,接下来包括的是计算步骤,这些步骤是以一步一步的提纲形式解释的。我们使用这种提纲式的方法是由于后面跟着的计算实例也是以同样的次序进行的,这样就使得后面的计算能很快地参照前面确切的计算步骤和方程。由于我们的目的是帮助学生们理解所包含的计算过程,所以这些计算实例也非常简单(利用人为的数据组)。在某些情况下,这种折衷办法既简单又带着生态统计的真实性。在这本教科书里,我们尽量点出,哪儿存在着这种折衷性,在这些示例的后边,我们进一步提供了使用计算机计算的一些其它例子。

在每一章方法的后面我们还提供了一节附加讨论和推荐的总结,附加的论题由简单的描述和选择的文献组成,其目的是强调一下在前面的介绍中没有包含到的一些方法。同学们会发现这些处理是不完备的,更确切地说,我们的目的是摘录一些问题和进展,这些也许能引起对于每种方法有个基本了解的同学们的兴趣。这些参考文献也为进一步学习提供了指南,最

后总结和推荐一节其目的是(强调)突出关于特定技术使用的重要结论。

在生态学中增加数量方法使用的重要因子是计算机的广泛普及。我们感觉计算机是提高学生们学习效果的一个重要教学工具,因而我们提供了一些 BASIC 计算机程序和教科书中的一些例子一起供同学们使用,学生们也可以用自己的数据。但我们还是鼓励学生们在应用这些程序之前利用书中一些示例操作一遍,以获得对所包含的一些方法的了解。这些计算机程序是专门为在 MSDOS/PCDOS 下,利用 BASIC 编译和编辑的微机上使用的。正如书中的一些例子,对于小型的数据组,微机是十分理想的,因为计算它们需要的时间比较少,对于大的数据组,对于我们所提到的方法有众多的统计程序可在大、中、小型计算机上使用。

最后,衷心地感谢新墨西哥州立大学、北卡罗莱那州立大学和圣地亚哥州立大学的许多学生鼓励我们撰写这本书;同时,感谢众多的同事们在阅读各章的草稿时所给予的巨大帮助,这些同事包括 Mike Austin, Peter Diggle Harvey Gold, Stuart Hurlbert, Dennis Knight, Robert Knox, Bob Peet, Fred Smeins 和 Brian Walker;也十分感谢我们家庭给予的支持、鼓励和忍耐。

John A. 拉德维格

James F. 蓝诺兹

1988 年 2 月

目次

第一部分 群落生态数据

1. 背景

- 1.1 实验与观测数据, 1
- 1.2 生态取样, 1
- 1.3 生态学数据和微机, 4
- 1.4 统计定义, 4
- 1.5 总结和推荐, 5

第二部分 空间格局分析

2. 背景

- 2.1 矩阵展示, 7
- 2.2 选择的文献, 9

3. 分布方法

- 3.1 基本方法, 10
- 3.2 步骤, 12
- 3.3 例子: 木蜂幼虫, 17
- 3.4 例子: 杨树上的织蛾, 21
- 3.5 例子: 苹果树叶上的螨, 22
- 3.6 分布模型的附加讨论, 23
- 3.7 总结和推荐, 24

4. 样方方差法

- 4.1 基本方法, 25
- 4.2 步骤, 26
- 4.3 举例: 计算, 29
- 4.4 例子: 荒漠灌丛中的麻雀巢, 31
- 4.5 例子: 草场上的火蚁, 32
- 4.6 样方方差法的附加讨论, 32
- 4.7 总结和推荐, 34

5. 距离方法

- 5.1 基本方法, 35
- 5.2 步骤, 35
- 5.3 举例: 计算, 37
- 5.4 例子: 松林, 38
- 5.5 例子: 荒漠中的收获蚁, 39
- 5.6 距离方法的附加讨论, 40
- 5.7 总结和推荐, 41

第三部分 种—多度关系

背景, 3

1. 背景, 1.3

2. 样方方差法, 2.3

3. 样方方差法, 3.3

4. 样方方差法, 4.3

5. 样方方差法, 5.3

6. 样方方差法, 6.3

7. 样方方差法, 7.3

8. 样方方差法, 8.3

9. 样方方差法, 9.3

10. 样方方差法, 10.3

11. 样方方差法, 11.3

12. 样方方差法, 12.3

13. 样方方差法, 13.3

14. 样方方差法, 14.3

15. 样方方差法, 15.3

16. 样方方差法, 16.3

17. 样方方差法, 17.3

18. 样方方差法, 18.3

19. 样方方差法, 19.3

背景, 3

1. 背景, 1.3

2. 样方方差法, 2.3

3. 样方方差法, 3.3

4. 样方方差法, 4.3

5. 样方方差法, 5.3

6. 样方方差法, 6.3

7. 样方方差法, 7.3

8. 样方方差法, 8.3

9. 样方方差法, 9.3

10. 样方方差法, 10.3

11. 样方方差法, 11.3

12. 样方方差法, 12.3

13. 样方方差法, 13.3

14. 样方方差法, 14.3

15. 样方方差法, 15.3

16. 样方方差法, 16.3

17. 样方方差法, 17.3

18. 样方方差法, 18.3

19. 样方方差法, 19.3

1

6

10

25

35

6. 背景

42

6.1 矩阵展示, 42

6.2 选择的文献, 43

7. 分布模型

44

7.1 基本方法, 44

7.2 步骤, 46

7.3 例子: 路中的昆虫, 48

7.4 例子: 荒漠群落的盖度数据, 50

7.5 例子: 纽约种禽, 51

7.6 对数正态分布模型的附加讨论, 51

7.7 总结和推荐, 53

8. 多样性指数

54

8.1 基本方法, 54

8.2 步骤, 54

8.3 例子: 荒漠中蜥蜴的多样性, 60

8.4 例子: 佛罗里达河口湾的鱼, 61

8.5 例子: 森林树木的多样性, 63

8.6 多样性的附加讨论, 64

8.7 总结和推荐, 66

第四部分 种的亲合性

9. 背景

9.1 矩阵展示, 67

9.2 选择的文献, 69

10. 生态龕重叠指数

10.1 基本方法, 70

10.2 步骤, 71

10.3 例子: 土蜂共同生存的计算, 74

10.4 例子: 鸟食谱重叠, 76

10.5 龕重叠问题的讨论, 77

10.6 总结和推荐, 78

11. 种间关联

11.1 基本方法, 80

11.2 步骤, 81

11.3 举例: 计算, 86

11.4 例子: 巴拿马蟑螂, 87

11.5 例子: 威斯康星森林, 89

11.6 种间关联的附加讨论, 90

11.7 总结和推荐, 93

12. 种间协变

12.1 基本方法, 94

94

- 12.2 步骤,94
- 12.3 举例: 计算,97
- 12.4 例子: 威斯康星森林,98
- 12.5 种间协变的附加讨论,99
- 12.6 总结和推荐,100

第五部分 群落分类

13. 背景

101

- 13.1 矩阵展示,102
- 13.2 选择的文献,104

14. 相似函数

105

- 14.1 基本方法,105
- 14.2 步骤,105
- 14.3 例子: 计算,109
- 14.4 距离函数的评价,110
- 14.5 例子: 巴拿马蟑螂,112
- 14.6 例子: 威斯康星森林,113
- 14.7 相似性函数的附加讨论,114
- 14.8 总结和推荐,115

15. 关联分析

116

- 15.1 基本方法,116
- 15.2 步骤,116
- 15.3 举例: 计算,118
- 15.4 例子: 威斯康星森林,119
- 15.5 关联分析的附加讨论,119
- 15.6 总结和推荐,120

16. 聚类分析

121

- 16.1 基本方法,121
- 16.2 步骤,123
- 16.3 举例: 计算,124
- 16.4 例子: 巴拿马蟑螂,125
- 16.5 例子: 威斯康星森林,126
- 16.6 聚类分析的附加讨论,129
- 16.7 总结和推荐,130

第六部分 群落排序

17. 背景

131

- 17.1 矩阵展示,132
- 17.2 选择的文献,133

18. 极点排序

134

- 18.1 基本方法,134
- 18.2 步骤,134

- 18.3 举例: 计算, 136
- 18.4 例子: 巴拿马蟑螂, 138
- 18.5 例子: 威斯康星森林, 139
- 18.6 极点排序的附加讨论, 140
- 18.7 总结和推荐, 141

19. 主分量分析 142

- 19.1 基本方法, 142
- 19.2 步骤, 142
- 19.3 举例: 计算, 145
- 19.4 例子: 巴拿马蟑螂, 151
- 19.5 例子: 威斯康星森林, 152
- 19.6 主分量排序的附加讨论, 153
- 19.7 总结和推荐, 154

20. 对应分析 156

- 20.1 基本方法, 156
- 20.2 步骤, 156
- 20.3 举例: 计算, 158
- 20.4 例子: 巴拿马蟑螂, 160
- 20.5 例子: 威斯康星森林, 161
- 20.6 对应分析的附加讨论, 163
- 20.7 总结和推荐, 164

21. 非线性排序 165

- 21.1 基本方法, 165
- 21.2 步骤, 166
- 21.3 举例: 计算, 169
- 21.4 例子: 巴拿马蟑螂, 172
- 21.5 例子: 威斯康星森林, 173
- 21.6 非线性排序的附加讨论, 175
- 21.7 总结和推荐, 176

第七部分: 群落的解释

22. 背景 178

- 22.1 矩阵展示, 178
- 22.2 选择的文献, 178

23. 分类的解释 180

- 23.1 基本方法, 180
- 23.2 步骤, 181
- 23.3 举例: 计算, 183
- 23.4 例子: 巴拿马蟑螂, 185
- 23.5 例子: 威斯康星森林, 186
- 23.6 分类解释的附加讨论, 188

23.7 总结和推荐,189

24. 排序的解释

190

24.1 基本方法,190

24.2 步骤,191

24.3 举例:计算,192

24.4 例子:巴拿马蟑螂,194

24.5 例子:威斯康星森林,195

24.6 排序解释的附加讨论,196

24.7 总结和推荐,198

参考文献

软盘程序(BASIC)说明

第一部分 群落生态数据

第一章 背景

统计生态学包含许多用来探索生物群落格局(Patterns)的数量方法,这些格局是多种多样的,包括:群落内部某一个种的空间分布和植物种间关系;群落之间植物种间关系。因此,统计生态学包含于大家所知道的数学或数量生态学这一广阔的领域内了。数量生态学包括对种群动态和群落格局的研究。

1.1 实验与观测数据

在群落生态学中,生态数据可被视为实验和观测方法的产物(Goodall 1970)。为了阐明各种类型生态研究的不同点,对于这些方法加以区别是有帮助的,因此,我们收集了数据的类型和限制。

一种实验方法是预先假定这个群落受实验操纵的,也就是说我们可以将一个群落分成几个重复的部分,在这些重复的部分上可以进行各种处理和对照实验,所以在测量反应方面检验出的任何不同都可归咎于实验处理。另一方面,使用观测的方法,我们可以对群落一系列的情况进行自然的测定而不是人为的测定。这就留给我们两个选择:(1)研究不同的样本,这些样本是在同一时间但不同的情况下获得的(如湖泊中心与边缘的取样);(2)研究的样本在同一地点但不同的时间(如:在夏季和冬季对湖心浮游植物的取样)。

群落生态学家通常对获得一个群落内大量变量的信息感兴趣,但对这些变量没作更多的处理,这也就是我们通常使用的观测方法,它是归纳性的、非实验性的和多变量的(Noy-meir 1970)。通过这本书的例子和主题的选择它将变得很明显,这也是我们在这个入门读物中将要使用的方法。

群落生态学的许多工作是在要阐明和描述数据组的格局的希望下得到推动的,而不是通常的对预先假设的检验(Green 1980),这是一个需要经常强调的非常重要的区别。在这本初级读物中我们考虑最多的是格局的检测方法(Pattern detection methods),正象 Greig-Smith (1971)指出的那样,群落生态学中的观测通常是通过空间或时间对可能目标的一些变化,包括如一个地区内所有种类组成的估测,种类成分和环境因子的相关和种格局在空间和时间上变异性的研究而获得的,在某些情况下通过样本对特定格局的检验可能导致关于生态群落潜在格局因果假设的形成(Noy-Meir 1970),这个假设可以通过进一步的工作得到检验(也可能用实验的方法)。

1.2 生态取样

观测研究的各个阶段如图 1.1 所示,第一步包括研究目的的确切定义。因为它影响数据收

集和分析等所有一系列的各个方面,这第一步限定了研究的领域和范围。

一个准确取样计划的选择也是十分重要的。正如 Greig-Smith(1983)阐述的那样,“数量数据的价值……依赖于用于获得这些数据取样过程”。任何一个计划的采用必须与研究的目的密切相关,而且应设计为能获得最大的信息量的方案,作为调查所付出的努力和时间的报答。尽管对于取样我们没有提供独立的特定指导,但在这本书里我们试图突出一些例子,在这些例子中取样的问题可能出现,同时在一些分析中也将强调取样的重要性。关于适合于群落生态学取样方法和理论的杰出的介绍,我们请同学们参见 1.5 节中列举的参考文献。

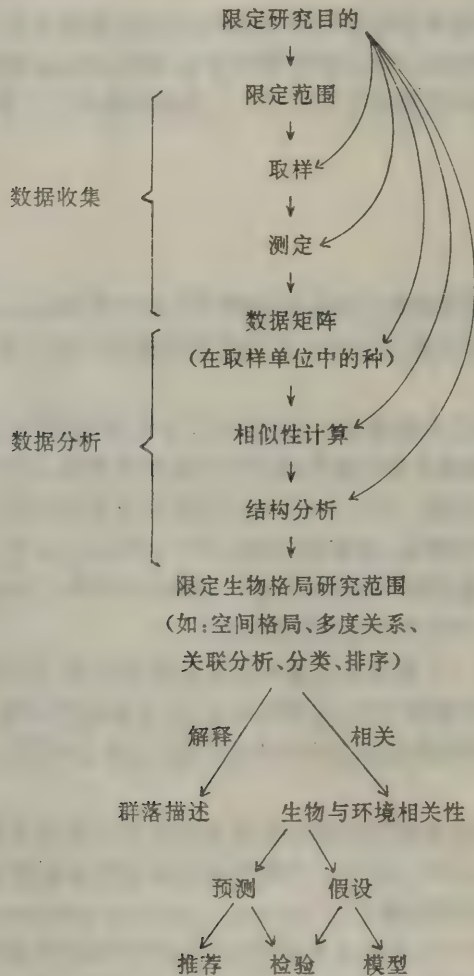


图 1.1 一个观测的生态学步骤(引自 Noy-Meir, 1970)

一个成功的取样计划包括取样单位(Sampling unit, SU)的确切选择,生态学中常规的取样方法包括样方、植物叶子、光陷阱、土钻、点陷阱、个体和样线法(表 1.1)。有些取样单位以自然形式出现(如植物叶子),另外一些是人为定义的(如样方),两者之间的重要区别在 2.1 节中讨论,在这本书里所提供的例子将会涉及各种 SUs。注意把取样单位称为样本(Samples)是不正确的,一个样本是由一个取样单位的集合构成(Fielou 1974)。例如,使用 1m^2 的取样单位取

样,通过这一研究区域的样本可能由 20 个随机放置的这种 SUs 组成。

一旦取样的过程和 SU 的选择确定了,在一个群落中所感兴趣的特定的测定(如出现—不出现、生物量、密度)就可以开始了,之后这些数据又被列成生态数据矩阵(Ecological data matrix),这是一种总结大量数据的方便的方法,也是我们进行分析的基本单位,这种数据矩阵是对每个 SU 测定结果的一种长方形显示,依研究目的的不同,有两种基本类型的数据矩阵,第 1 种是与时间动态(Temporal dynamics)有关的研究(如群落演替),这种数据矩阵表示的是对种通过时间的测量(图 1. 2a),时间的间隔依赖于研究的特定目的,环境因子(如土壤水分含量、土壤 pH、土壤温度)在这种取样单位中常常也被同时收集,第二种类型的数据矩阵是处理通

表 1. 1 在生态学各领域使用的 SUs 和典型变量测定的例子

领域	测量参数	取样单位
种群生态学	多度参数(如密度)	陷阱线、网格、平行样线
生理生态学	有机体反应(如光合成)	个体动物、植物或微生物
行为生态学	反应参数(如喂养速率)	个体动物、植物或微生物
微生物生态学	生长/腐烂参数(如凋落物损失)	培养皿或凋落物包
群落生态学	种多度或存在/不存在	样地、样方或样线



图 1. 2 2 种基本的生态数据矩阵:(a)通过时间在某一地点对种和环境因子的测量(利用这种 SUs 共观测 t 次)和(b)通过空间用 N 个 SUs 对种和环境因子测定数据,也就是在景观的不同位置的测定。

过空间(Space)分布而获得的取样单位的测量值的(图 1. 2b)。通过实验设计使 SUs 实际的空间分布得到测定(如随机放置样方)。当然,通过空间的研究也可以利用季节的或年度的重复来实现,例如,通过时间对空间动态的研究。在本书每一部分开始的一章里,我们将会给出分析所用的生态数据的特定形式。

1.3 生态学数据和微机

这本书所提供的程序是适合微型计算机的,设计这些程序的目的是帮助学生理解本书的各个例子和便于他们用自己的数据进行实习。同大型计算机相比,微机有可利用的内存和计算速度的限制,所以这些程序不适用于通常群落生态学家所研究的大量的数据。处理大量的数据,大型计算机当然是既有效又节省时间。在这本入门读物中所包含的许多统计技术是适合于生态学家在大型计算机的软件包上利用的(见 1.5 节)。

在任何类型的计算机上,当使用现成的程序时,众多的错误就潜在地存在着,使用者可以结合逐个地按键作各种类型分析,这需要一些有关计算和假设的知识,否则将导致方法的不确切应用和对结果的错误解释,同时也许有计算上和逻辑上的错误用检验数据没检验出,当用其它的数据(包括你自己的)时,将产生一个不正确的结果。例如,在我们这些程序的早一版本中,只要是生态数据矩阵的组成行(种)大于列 SUs 就能获得正确的结果。也凑巧,我们用的所有检验数据在发展这些程序时都是标准的行大于列。当然,当第 1 个学生使用这个程序处理恰巧行少于列的数据时,错误就出现了。

当我们试图解释微机输出的结果时,舍入的误差也许存在,计算机的存贮和计算都是有固定位数的,最大值随着计算机的大小不同而不同,因此在不同容量的计算机上使用同样的程序可能产生不同的结果,可以归咎于这类舍入误差,同学应该注意这些可能性,在计算机上舍入误差的深入讨论见 Nelder(1975),Lewis 和 Doerr(1976)。

1.4 统计定义

本书经常使用的一些术语定义如下:

1. 种群:在一个生态群落中可以观察到的个体 SUs 全部的集合,从这个种群可以取得一个样本并作出统计推断。
2. 参数:种群的一些已知特征(如密度)。
3. 统计量:种群参数的一个估测值(如通过对某个种群取样获得的种群密度的估测值)
4. 准确性:使用一定的取样步骤获得的对种群参数真实值统计量接近程度的测量。
5. 精确性:使用一定的取样过程在重复的样本中统计量重复性程度的测量,标准误被认为是取样精确性的基本表示。注意,没有准确性也是可能获得精确估计的,因为除了有偏差的,特殊的取样步骤也可能是很精确的。
6. 偏差:由一个具体统计量对真实的种群参数一致的过高或过低的估计。统计偏差通常是由于一个取样步骤,或在某些情况下,由于一个公式所造成的一致不准确的结果。
7. 误差:种群参数的统计量和真实值间的差异,认识到偏差和误差也许均是由原始数据的收集和对数据一系列的处理所造成的,是十分重要的。

1.5 总结和推荐:

1. 群落生态学的许多工作是在阐明和描述格局的意愿下得到推动的,而不是在通常对假设的检验得到推动的,这是需要记在心中的一个重要的区别(1.1节)

2. 任何研究的第一步就是明确限定它的研究目的。

3. 设计某一领域的研究(或解释生版的一些东西)最重要的一步是确定确切的取样单位(1.2节,表1.1),使用这些SUs获得样本。

4. 关于取样所参照的一些重要参考文献包括 Greig-Smith(1983)所著的第二章, Southwood(1978)所著的第二章, Green(1979)的第二、三章, Cochran(1963), Williams(1976), Eberhardt(1976), Cox(1985)一到四章, Myers 和 Shelton(1980)的附录1-9, Mueller-Dombois(1974)第四章和 Knapp(1984)。

5. 计算机是分析数据的一种有力的工具,但也要注意使用现成的程序时存在的一些潜在的错误(1.3节),本书所提供的绝大多数分析可以在大型计算机的软件包上进行,这些软件包括 BMDP(Dixon 和 Brown 1979), ORDIFLEX(Gauch 1977), SAS(Ray 1982), NT/SYS(Rohlf 等 1971), CLSTAR/CLUSTID(Romesburg 1984), CLUSTAN(Wishart 1969)以及 Cooley 和 Lohnes(1971), Orloci(1978, 1985)和 Williams(1976)书中的一些程序。Green(1979), Legendre 和 Legendre(1983)对这些软件包进行了评论。

第二部分 空间格局分析

第二章 背景

植物和动物的空间格局是生态群落学中的一个重要的特征,这是我们在观察任何群落时最先看到的现象之一,也是一群生活的有机体的最基本的成分之一(Connell 1963)。

在群落中三种基本类型的格局得到承认:随机的(Random)、集聚的(Clumped)和均匀的(Uniform)。在本书的这一部分里我们将提供检验空间格局的各种方法。一旦一种格局被确定,生态学家就可以提出并且检验假设以解释引起这种格局的潜在因子,因此查清空间格局的最终目的是产生与生态群落有关的假设(Williams 1976)。例如,这是 George 和 Edwards (1976)研究的一个情况,他们调查了威尔士的一个水库中风诱导水运动对浮游植物和藻类非随机的、水平格局形成的重要性;另一个研究也是在威尔士,Doncaster(1981)发现在一个大的岛屿上蚂蚁穴的聚集格局很大程度上受暴露的坡度水分和兔子对该岛屿植被啃食的影响。最后一个例子是:Lamont 和 Fox(1981)在西澳大利亚研究了金合欢树两个强度水平的空间格局,即树的聚块(Clump)之间和聚块内部的,他们发现两个水平的格局均受干旱和动物选择性啃食的影响。

下面的一些机理经常被用来解释在群落中观察到的格局(Pemberton 和 Frey 1984),一个有机体种群的随机格局意味着环境的同类性和/或非选择性的形为格局,另一方面,非随机性的格局(集聚的和均匀的)意味着一些因素迫使该种群以这种方式存在。集聚意味着个体聚集在生境中相对适宜的部分,这也是由于它们喜欢群居的行为和环境的异质性以及再生方式等原因。均匀分布是由于个体间负的相互作用,如对食物或空间的竞争(Competition)所造成的,检验一个格局和分析其产生的可能原因是独立的问题。进一步因应该记在心中的是:自然是多因子的;许多相互作用过程(生物的和非生物的)对格局的存在有重要作用(Quinn 和 Dunham 1983)。

Hutchison(1953)是最早考虑到群落空间格局的重要性并确定引起有机体形成格局的各种因子的生态学家之一。他指出格局的形成原因主要有:(1)由于外部环境力量(如风、水流、光强度)作用而产生的水平因子的作用;(2)可归咎于有机体再生方式的再生因子(如克隆和后裔的再生方式);(3)由于先天行为的群居因子(如领域行为);(4)由于种内相互作用(如竞争)产生的相互作用因子;(5)由于前面几个因子随机的任何差异而产生的随机因子。这样空间格局可归咎于种内在的(如再生、群居和相互作用)或外部的(如水平作用过程)作用。Kershaw(1973)、Southwood(1978)和 Greig-Smith(1979,1983)对种格局产生的原因作了进一步的讨论。

植物群落生态学家对许多空间格局分析(Spatial pattern analysis,简称SPA)方法的发展起了很大的作用,这可能是由于在植物群落中植物直立的特征和格局比较明显的特性。SPA通常限定在一个群落内小规模格局范围内使用,当然格局规模的选择是重要的。某个群落中,

各种程度种集聚出现在中等规模上,在一个规模上格局可以被检测出而在其它规模上却不能,事实上,在群落中格局检验的 SPA 技术是以在各种规模上探索格局为基础的。Pielou(1979)给出了一个用 SPA 对大规模生物地理研究应用的一个例子。

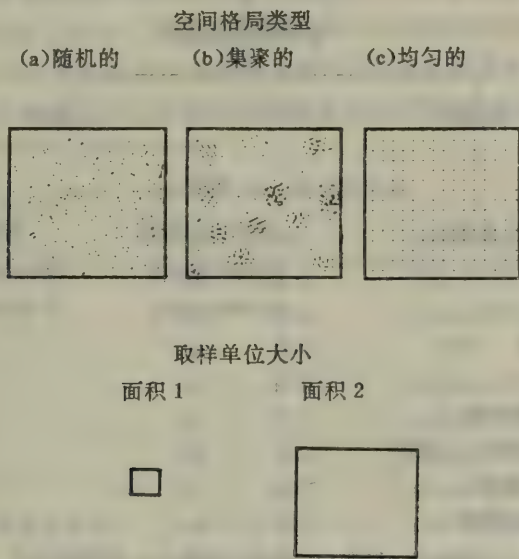


图 2.1 三种类型的空间格局:(a)随机的,在随机的格局中所有个体之间独立地分布;(b)集聚的,在这种格局中个体趋向于以成丛形式居住在一起;(c)均匀的,在这种格局中个体规则地占有空间。给出了两个取样单位大小

我们提供的 SPA 方法仅仅用于处理空间格局的检验,但通过不同季节和年度即通过时间去探索格局,即重复格局分析的研究也是很普遍的。

2.1 矩阵展示

SPA 模型是以通过取样单位(SUs)某个种的多度测定的数据(第 3、4 章)或群落内个体间的距离(第 5 章)为基础的。通过取样单位种类矩阵的显示如图 2.2。当解释 SPA 的研究结果

		取样单位										
		1	2	3	4	5	N	
种	a	■	■	■	■	■	■	■	■	■	■	
	b											
	c											
	...											
	s											
环境因子	w											
	x											
	...											
	z											

图 2.2 阴影部分表示的是 SPA 生态数据矩阵形式,在这里感兴趣的是通过许多(N)SUs 单一种(如种 a)个体的分布,取样单位既可是自然的(如松树果球),也可是人为的(如样方)

时,自然的取样单位具有重要的含义,这些含义引起了一些问题(第 3、4 章所描述的)。由于这些问题,使得利用距离的 SPA 技术具有一定的优点,这些优点我们将在第 5 章说明。

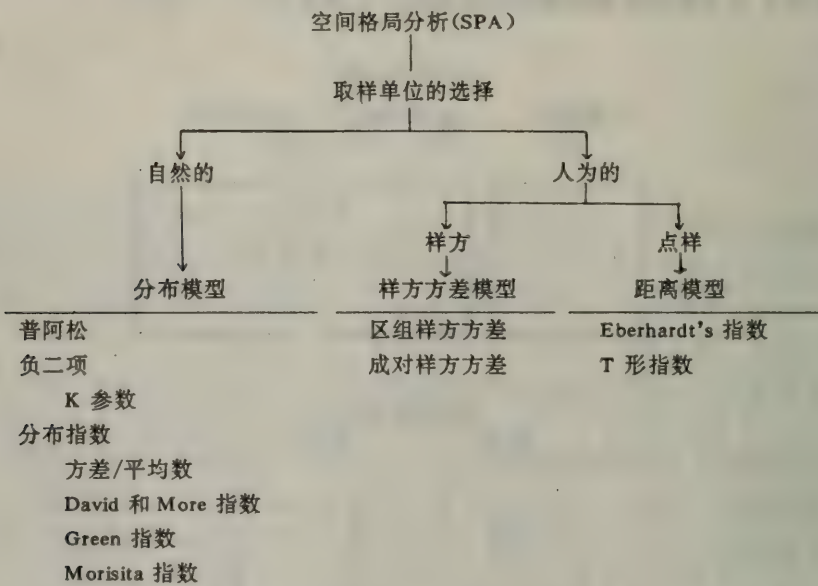


图 2.3 根据 SU 选择的 SPA 模型的类型

Pielou (1977, 1979) 区别了自然和人为的取样单位, 我们可以定义自然 SUs 为: 有机体出现在一个生境的不连续片断的 SUs。例如, 腐烂园木内的白蚁, 水果中发现的虫子, 粪块中的分解者, 这些园木、水果、粪块被认为是自然的 SUs, 以从这些自然的 SU 获得的有机体多度为基础的各种统计量和指数可能转化为关于空间格局有意义的信息 (Pielou 1979)。

对于出现在连续生境的有机体。如森林中的树、湖泊中的浮游动物、普列利(高草草原)上的草, 应用一些人造的取样单位(如小区或样方)获得样本是必要的。这样, 一个样本的具体统计量, 如每个样方中树的平均数目在表达关于空间格局有意义的信息方面是很少有帮助的, 这是因为不同大小的取样单位, 可能导致不同的结论, 如图 2.1 所示。由该图可以看出, 用面积 1 的 SU 对集聚种群重复的随机取样将会获得许多具有 0 个或少数几个个体的 SUs, 偶尔也可以获得具有许多个体的 SUs, 这将表明该种群是集聚的, 但当我们使用面积 2 的 SU 时, 将会趋向于得出每个取样单位个体数目大致相同的结果, 这又表明该种群的格局是均匀的而不是集聚的。由这个例子可以看出, 在连续生境中的格局检验将依赖于与 SU 大小有关的格局规模, 问题是这种关系在我们取样之前未必知道。有关取样、取样单位和空间格局的关系的进一步讨论我们推荐 Pielou (1977) 的著作。

2.2 选择的文献

在表 2.1 中我们给出了所选择的一些 SPA 生态应用的一些例子, 根据选择 SU 的不同, 格局检验的研究可以分别地被放在 SPA 模型中三个性质不同的领域, 这些已在图 2.3 中列

出。在第三章提供的分布模型(Distribution models)是以通过一自然取样样本种多度数据的相对频率分布的观察值为基础的。一些用来测定集聚程度的分布指数也是以相对频度分布为基础的,第四章提供的样方方差模型(Quadrat variance models)适用于当观测值是来自人为的取样单位时。第五章阐述的距离模型(Distance models)适用于当观察值是来自点和个体间的距离时。有些生态学家尽管知道正如 2.1 节讨论的那样 SPA 结果和结论依赖于 SU 的大小和形状,在处理用人为的 SU 获得的数据时仍旧使用分布模型,同学们阅读完第三章就会发现这是不正确的原因了。我们附带地提一下这些类型的错误正是我们想要避免的。

表 2.1 选择有关 SPA 模型的文献

地点	群落	方法 ^a	SU ^b	参考文献
密西根	森林	DS	点样	Dice 1952
密西根	森林	QV	网格	Spires 和 Klosterman
		DS	点样	1981
英格兰	草地	DU	样方	Clapham 1936
澳大利亚	盐沼泽	QV	样方	Goodall 1963
加利福尼亚	荒漠	DS	点样	Gulmon 和 Mooney 1977
威尔士	湖泊	DU	瓶子	George 和 Edwards 1979
加拿大	禾草—豆科牧草草场	QV	样方	Lamont 和 Fox 1981
澳大利亚	林地	QV	网格	Turkington 和 Cavers 1979
哥斯达尼加	森林	QV	网格	Hubbell 1979
实验室	浮游动物海藻	DU	养分斑块	Lehman 和 Scavia 1982
安第斯山脉	火裂鸟	DU	湖泊	Hurlbert 和 Keith 1979
威尔士	蚂蚁	DU	样方	Doncaster 1981
加拿大	地衣	QV	样线	Yarranton 和 Green 1966
亚南极	苔藓泥炭藓	QV	样方	Usher 1983

^a:DU=分布模型;QV=样方方差模型;和 DS=距离模型

^b:SU 注意在每个研究中使用的取样单位类型

第三章 分布方法

在这一章里我们将描述在一个群落中用来检验和测量种空间格局的统计分布(Statistical distribution)和分布指数(Indices of dispersion)的使用。如果一个种的个体通过不连续的空间分布,如植物叶片上的介壳虫,而且样本是来自每个SU 个体的数目,那么就可能利用频度分布的形式来总结这些数据,这个频度分布是由含有0 个个体、1 个个体、2 个个体等等取样单位的数目组成。这种基本的数据组成是以后我们将要描述的格局检验方法(Pattern detection methods)中要使用的形式。注意所研究的种是被视为将出现在不连续地点或自然的取样单位中,如叶子、水果、树(见 2.1 节)。

3.1 基本方法

每个取样单位个体数目的观察频度分布能够告诉我们关于空间格局的一些什么呢? 举个例子,考虑到某一个植物叶片上介壳虫可能分布的一些方式。如图 3.1 所示,总数为 30 的昆虫分别以随机的、集聚的和均匀的方式分配在 3 株植物的 10 片叶子上,当然如不考虑它分布的方式,每株植物的每片叶子上(SU)的昆虫数目将为 3。

在随机分布的情况下,每片叶子具有相同的机会被昆虫侵入,同时在一叶片子上有昆虫出现并不影响其它个体出现与否,对于随机格局来说这种频度分布以每片叶子上有 3 个个体为峰值(平均数),其它数目的分布在峰值的两侧。在集聚分布的格局中,多数叶片上没有昆虫个体而少数叶片上有许多个体,均匀分布格局与之相比,其绝大多数叶子具有 3 个个体。

同时注意每片叶子上个体数目的方差是如何受昆虫分布格局影响的。在随机格局中(植物 A),方差(2.6)接近于平均数,而在个体集聚或集聚分布在少数叶片上(植物 B),方差(18.2)远远大于平均数;在个体分布呈均匀的情况下(植物 C),方差(0.22)远远地小于平均数(当然,如果每片叶子上昆虫个体的数目完全相同,方差将为 0)

总之,每个取样单位个体数目的平均数和方差之间的关系受种群分布潜在格局的影响。我们现在可以定义三种基本类型的格局和它们方差与平均数之间的关系,在这里 σ^2 = 方差, μ 代表平均数。

1. 随机格局: $\sigma^2 = \mu$
2. 集聚格局: $\sigma^2 > \mu$
3. 均匀格局: $\sigma^2 < \mu$

由于它们方差与平均数成分有差异,故有一定的统计频度分布被用来作为这些类型生态格局的模型。

1. 泊阿松分布($\sigma^2 = \mu$)对应于随机格局
2. 负二项分布($\sigma^2 > \mu$)对应于集聚格局
3. 正二项分布($\sigma^2 < \mu$)适用于均匀分布。

这三种统计模型经常被用在空间格局的研究上,同时我们也应当认识到其它统计分布也许能应用到这种研究上(Pielou 1977)。

在群落生态学中格局检验最初一步往往是对一个假设的检验,这个假设就是每个取样单位个体数目的分布是随机的,如果这个假设被拒绝,那么这个分布就倾向于集聚的(通常)或均匀的(很少),如果是倾向于集聚分布,与负二项分布的一致性就可以被测定,同时以方差与平均数的比为基础的一些分布指数就可以用来测定集聚的程度,因为在生态群落中均匀格局相对稀少,这种情况我们这里就不考虑了,对此感兴趣的,参见 Elliott(1973)给出的均匀格局的一些生态例子。

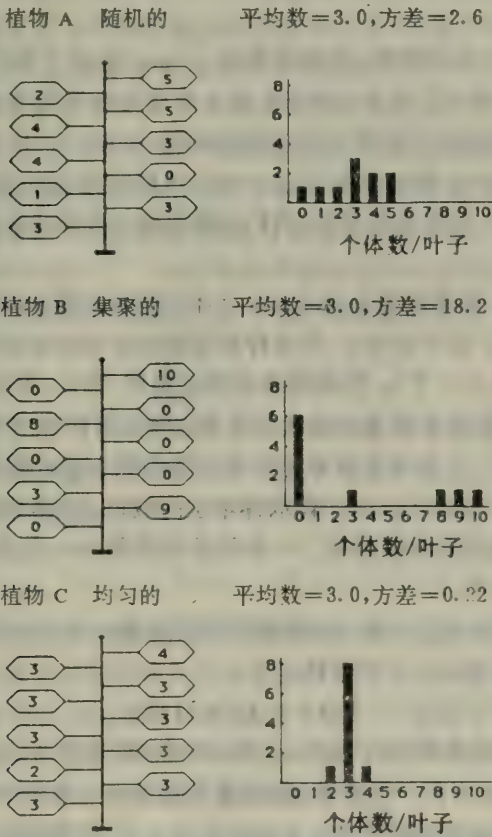


图 3.1 30 个介壳虫在 10 片叶子上(自然 SU)可能的分布形式和每片叶子昆虫的数目,同时给出了植物 A(随机的)、B(集聚的)、C(均匀的)的每片叶子上昆虫数目的平均数和方差

具体进行之前,我们应当注意的几点是:第一,拒绝随机假设失败后,只意味着我们在应用手头上一些种的特定数据去检验非随机性的格局不成立,在同样一个群落中一个独立收集的数据也将会导致同样的结论吗? 第二,我们应当有理由地提出假设(Pielou 1974):一个假设应该是能站得住脚的而且应以大家共同的感觉和生物学知识的混合体为基础。这第二点有一个重要的分支是与第一点有关的。使观察的频度分布与理论统计分布(如泊阿松过程)相象(也就是两者间有一个统计的一致性)是很平常的,即使这个理论模型的假设不能被这组数据满足。

事实上,我们可能接受了一个没有正当生物学理由的零假设。第三,我们不应把我们的结论单单建立在显著程度检验的基础,所有可利用的信息源(生态的和统计的)都应有机地利用起来。例如,建立在一个小样本上的零假设没有被拒绝应该被视为是对这个假设的薄弱肯定(Snedecor 和 Cochran 1973)。最后,记住检验空间格局(我们这里的目标)和解释它们产生的可能因子是独立的问题(见第 2 章)。

3.2 步骤

3.2.1 泊阿松概率分布

对于一随机分布的有机体种群,泊阿松模型($\sigma^2 = \mu$)给出了每个取样单位个体数目的概率,提供了应遵照以下条件:(1)每个自然 SU 某个个体的出现具有相同的概率,(2)在一个取样单位中某一个体的出现与否不影响其它个体的出现,(3)每个 SU 是同等有效的,(4)每个取样单位中个体的数目相对于能够出现在这个 SU 可能性的最大值是比较低的(见 Greig-Smith 1983, pp. 58—59)。计算这些概率,我们仅需要每个取样单位个体的估测值(μ),这个过程的大致步骤如下:

步骤 1. 提出假设。该假设就是每个 SU 个体的数目是遵从泊阿松分布的。如果该假设没被拒绝,那么我们就可下这样的结论,即该种群是随机分布的或者事实上该种群是非随机分布的,但没被检验出来(见 3.1 节)。如果该假设遭到拒绝,那么,可能是非随机的格局,其它的 SPA 模型也许可以被利用(其它的可能性我们将在这章后面的部分中讨论)。

步骤 2. 频度分布, F_x 。由每个取样单位中个体的数目组成的样本数据被总结为频度的分布。也就是含有 $x=0, 1, 2, \dots, r$ 个个体取样单位的数目。这里应该注意的是相当大的样本规模是很有必要以这种形式安排数据的。一条经验规律是 SUs 的最小数目(N)应大于 30,如果小于这个数字,见 3.23 节。

步骤 3. 泊阿松概率, $P(x)$ 。在一个取样单位中发现 x 个个体的概率即 $P(x)$ 。由泊阿松模型给出,在这里 $x=0, 1, 2, \dots, r$ 个个体。

$$P(x) = (\mu^x e^{-\mu}) / x! \quad (3.1)$$

在这儿 e 为自然对数的底数(2.7183), $x!$ 为 x 的阶乘[例如 $x=3, x! = (3)(2)(1)=6$]。平均数(μ)是泊阿松模型中仅有的一个参数。统计量 \bar{x} 是由作为每个 SU 个体数目的平均数计算出来的。利用作为 μ 的估计值,每个 SU $x=0, 1, 2, \dots, r$ 的概率是

$$\begin{aligned} P(0) &= e^{-\bar{x}} \\ P(1) &= (\bar{x})^1 e^{-\bar{x}} / 1! \text{ 或 } (\bar{x}/1)P(0) \\ P(2) &= (\bar{x})^2 e^{-\bar{x}} / 2! \text{ 或 } (\bar{x}/2)P(1) \\ &\dots \dots \\ P(r) &= (\bar{x})^r e^{-\bar{x}} / r! \text{ 或 } (\bar{x}/r)P(r-1) \end{aligned}$$

由于这些值是概率,故其总和为 1,注意一旦 $P(0)$ 被计算出来,这些概率简单化的计算形式省略了计算阶乘的必要(见上面方程右边的形式)

步骤 4. 泊阿松频度期望值, E_x 。这个泊阿松模型是个概率分布,当每个概率与这个样本

中 SU_s 总数 (N) 相乘时, 含有 $0, 1, 2, \dots, r$ 个体 SU_s 的期望数就能被确定。因此, 使 E_x 代表每个 SU 中 $x=0, 1, 2, \dots, r$ 个个体的期望频度, 我

$$E_0 = (N)P(0)$$

$$E_1 = (N)P(1)$$

$$E_2 = (N)P(2)$$

$$\dots \dots$$

$$E_r = (N)P(r)$$

们使每一个方程代表一个频度级, 就产生了一个总数为 $q=r+1$ 期望个体的频度级, 如果在分布的尾部频度下降的太低以致于个体的期望数目变得非常小, 这个频度级应分别使用下列规则 (Sokal 和 Rohlf 1981): (1) 如果 $q < 5$ 那么最低的期望值应不少于 5 个个体或 (2) 如果 $q \geq 5$, 那么最低的期望值应不少于 3 个个体。一个普遍原则是不能有小于 1 的期望值 (E_x), 或者更严格的, 不小于 5 (Poole 1974)。这些并不是严格的规则而且最小的期望值 1 和 3 也可以试一试。在频度级的划分可能明显影响检验结果的情况下, 我们推荐比较保守的数值。频度经过必要的分摊后频度级数还有残余, 这时频度级的数目 q 就是这个新的值。

步骤 5. 吻合性检验统计量, χ^2 。卡方吻合性检验是用来弄清观察的频度 (F_x , 步骤 2) 与期望频度 (E_x , 步骤 4) 相比吻合到什么程度, 这个卡方检验统计量按下式计算。

$$\chi^2 = \sum_{x=0}^q [(F_x - E_x)^2 / E_x] \quad (3. 2)$$

将这个统计量与在自由度为 $q-2$ 的 (回忆下 $q=r+1$, 步骤 4) 卡方概率表相比较, 如果这个卡方检验统计量在某个概率选择的水平上 (如 5%) 大于卡方表中的值, 那么我们就可以得出这样的结论, 即该频度分布是不可能遵从泊阿松分布的, 因此拒绝零假设。

3. 2. 2 负二项分布

负二项模型可能是应用在集聚的或通常所说的感染种群或聚集种群中的最普遍的概率分布 (Sokal 和 Rohlf 1981), 当与使用泊阿松模型有关的两个条件不能满足时 (3. 2. 1 节) 也就是条件 1 (个体在每个取样单位出现的概率相同) 条件 2 (在一个取样单位中某一个体的出现不影响其它个体的出现), 这就常常导致每个取样单位个体的数目具有较高的方差与平均数比, 按照前面所表示的那样, 这将表明一个集聚的格局可能存在。

负二项分布有两个参数: (1) μ , 每个取样单位个体的平均数目和 (2) k , 与集聚程度有关的参数。观测频度分布与负二项分布的一致性检验的步骤概括如下; 步骤与泊阿松分布描述的相似, 可以互相参照避免重复。

步骤 1. 提出假设。被检验的假设就是每个取样单位个体的数目遵循负二项分布, 因此存在一个非随机的或集聚的格局。拒绝这个假设失败, 这个生态学家就可能有个好的经验模型来描述这一观察的频度数据; 这并没有解释引起这种格局潜在的因子。Bliss 和 Calhoun (1954) 讨论了一些自然发生的现象, 这些现象可能引起有机体种群分布呈负二项分布。例如, 如果一个昆虫种群随机产的卵聚集在植物的叶片上 (泊阿松过程), 而且如果每群孵化的幼虫数目是独立地呈对数分布, 那么每片叶子幼虫的频度计数将遵循负二项概率分布。当然这个方案也许是可能的, Blish 和 Fisher (1953), Pielou (1977), 和 Solomon (1979) 给出了一些与可能的原因机

制(如前面的方案)有关的矛盾假设是怎样可以导致同样概率分布的例子。这点当使用概率模型时我们应常常记在心里。我们不应该仅仅以格局检验的方法为基础去探讨产生格局的原因。

步骤 2. 频度分布, F_x 。如 3.2.1 节, 每个取样单位个体的数目总结为概率分布的形式, 也就是含有 $0, 1, 2, \dots, r$ 个个体取样单位的数目。

步骤 3. 负二项分布概率 $P(x)$ 。在一个 SU 中发现有 x 个个体的概率也就是 $P(x)$, 这里的 $x=0, 1, 2, \dots, r, P(x)$ 由下式给出

$$P(x) = [\mu/(\mu + k)]^r \{ (k + x - 1)! / [x! (k - 1)!] \} [1 + (\mu/k)]^{-1} \quad (3.3)$$

参数 μ 由样本平均数 (\bar{x}) 估计, 参数 k 是集聚程度的测量, 在最大集聚程度时它趋向于 0, (见 3.6 节) 使用下面反复迭代方程获得 k 的估计值 (也就是 \hat{k})。

$$\log_{10}(N/N_0) = \hat{k} \log_{10}[1 + (\bar{x}/\hat{k})] \quad (3.4)$$

这里 N 是这个样本中 SUs 的数目。 N_0 是含有 0 个个体的 SUs 数目。首先将 k 的一个最初估计值代入方程的右侧 (RHS), 并且将获得的值与方程左侧 (LHS) 相比较, 如果 RHS 低于 LHS, 那么再试下个较高一点的 \hat{k} 值, 并且再比较方程的两侧, 以重复的方式继续这个过程 (确切选择较高或较低的 \hat{k} 值) 直到方程右侧的值与左侧的相等。第一次代换的 \hat{k} 值的一个较好的估计量由下式获得:

$$\hat{k} = \frac{\bar{x}^2}{s^2 - \bar{x}} \quad (3.5)$$

s^2 为样本方差的估计值。

当平均数比较小 (小于 4), 公式 3.4 是估计 \hat{k} 值的一种有效方法。如果平均数较大 (大于 4), 只有当所研究的种群聚集程度较强时, 这个反复迭代的方法才是有效的。如果某个种群的平均数 (\bar{x}) 和 (\hat{k}) 值均大于 4, 公式 3.5 同公式 3.4 相比, 公式 3.5 更适合估计 \hat{k} (集聚参数, 由公式 3.5 计算) 值。

一旦两个统计量 \bar{x} 和 \hat{k} 都得到了, 在一个取样单位中发现 x 个个体的概率, 也就是 $P(x)$, 即可按公式 3.3 计算如下: 这里 $x=0, 1, 2, \dots, r$ 个个体。

$$P(0) = [\bar{x}/(\bar{x} + \hat{k})]^r [(\hat{k} + 0 - 1)! / (0! (\hat{k} - 1)!)] [1 + (\bar{x}/\hat{k})]^{-1} \\ = [1 + (\bar{x}/\hat{k})]^{-1}$$

$$P(1) = [\bar{x}/(\bar{x} + \hat{k})]^r [(\hat{k} + 1 - 1)! / (1! (\hat{k} - 1)!)] [1 + (\bar{x}/\hat{k})]^{-1} \\ = [\bar{x}/(\bar{x} + \hat{k})] (\hat{k}/1) P(0)$$

$$P(2) = [\bar{x}/(\bar{x} + \hat{k})]^r [(\hat{k} + 2 - 1)! / (2! (\hat{k} - 1)!)] [1 + (\bar{x}/\hat{k})]^{-1} \\ = [\bar{x}/(\bar{x} + \hat{k})] [(\hat{k} + 1)/2] P(1)$$

...

$$P(r) = [\bar{x}/(\bar{x} + \hat{k})]^r [(\hat{k} + r - 1)! / (r! (\hat{k} - 1)!)] [1 + (\bar{x}/\hat{k})]^{-1} \\ = [\bar{x}/(\bar{x} + \hat{k})] [(\hat{k} + r - 1)/r] P(r - 1)$$

步骤 4. 负二项频度期望值, E_x 。如 3.2.1 节泊阿松模型计算的那样, 包含 x 个个体取样单位期望值由负二项分布的概率与样本中 SU 总和相乘获得。频度等级的数目 q 也按着在泊阿松模型所描述的那样计算。

步骤 5. 吻合性检验统计量, χ^2 。这个卡方检验统计量 \bar{x} 由公式 3.2 计算并且与自由度为 $q-2$ (参数) -1 也就是 $q-3$ 的卡方概率表中的值比较。

3.2.3 分布指数

正如 3.2.1 节中所描述的那样,理论的泊阿松分布方差与平均数相等。基于这一点,以方差与平均数比的一些指数被提了出来用以检验,(1)在泊阿松系列中方差与平均数相等性,(2)测定有机体种群的集聚程度。在这一节中我们提供了 3 个这样指数:分布指数(Index of dispersion),集聚指数(Index of clumping)和格林指数(Green's index)。

具体介绍之前,我们提出两点。首先,在生态学文献中(象在这章)方差与平均数比通常是指分布指数。但是,经过这些年,这个比例(和其它的)的一些变体被提了出来去测定集聚的程度,这些通称“分布指数”(包括集聚指数和格林指数)。其次,回忆一下我们考虑空间格局的问题时是以每个取样单位个体的数目为基础的,这里的 SUs 是不连续的,自然出现的统一体(2.1 节),在第 5 章我们将提供另一个分布指数,它是以有机体间的距离为基础的。

指数 1:分布指数。这个方差与平均数的比或分布指数(ID)是

$$ID = \frac{s^2}{\bar{x}} \tag{3.6}$$

同以前一样,这里的 \bar{x} 和 s^2 代表平均数和方差。如果这个样本与理论泊阿松系列一致,我们将希望这个比率为 1。所以利用下面的卡方检验统计量我们可以检验 ID 值背离 1 的显著性。

$$\chi^2 = \left(\sum_{i=1}^N (x_i - \bar{x})^2 / \bar{x} \right) \tag{3.7a}$$

$$= ID (N - 1) \tag{3.7b}$$

这里 x_i 是在第 i 个取样单位中个体的数目, N 是 SUs 总数,对于较小的样本规模($N < 30$), χ^2 是在 $N-1$ 个自由度下卡方的一个较好近似(关于这一点;注意公式 3.7a 与公式 3.2 的相似性,如果这个 χ^2 值在 0.975 和 0.025 的概率水平上($P > 0.05$)落在卡方值表之中,与泊阿松(随机)分布的一致性是可以接受的(即 $s^2 = \bar{x}$)。另一方面,如果 χ^2 值小于 0.975 概率水平上的卡方表中的值,表明格局为规则格局(即 $s^2 < \bar{x}$),而当 χ^2 值大于 0.025 概率水平的卡方表中的值时,表明是个集聚格局。在表 3.1 给出了图 3.1 中($N=10$)每种规模的昆虫分布的 ID 值和有关的 χ^2 检验统计量。每种情况,检验都肯定了潜在的格局。例如,在植物 A 中 ID 的值为 0.87,等于 1 的假设是可以接受的(即 $\chi^2=7.8$ 是在卡方表中 2.7~19.0 范围内的值)。

对于大的样本规模($N \geq 30$),($\sqrt{2\chi^2}$)倾向于正态分布,同时其它的检验统计量 d 可以由下式计算:

$$d = \sqrt{2\chi^2} - \sqrt{2(N-1)-1} \tag{3.8}$$

如果 $|d| < 1.96$,与泊阿松(随机分布)一致是可以接受的($P > 0.05$)。如果 $d < -1.96$,值得怀疑为规则分布,如果 $d > 1.96$,很可能是集聚分布(Elliott 1973)。当 N 较大(> 30),也应该进行对泊阿松分布的吻合性检验(见 3.2.1 节),并将这个结果与前面的方差与平均数比检验相比较。当然两个检验使用的是同样的数据,所以不是独立的。Greig-Smith (1983)讨论了这些情况,这两种检验之一可能检验出是明显的非随机性,另一个可能没有检验出是非随机性的,这再一次表明,将结论建立在单一的观测数据组和单一的检验是不妥的。

表 3.1 图 3.1 中的介壳虫种群分布指数(ID)和格林指数(GI), χ^2 统计量[公式(3.7)]
和 0.975 和 0.025 概率水平上的卡方临界值

植物	分布	ID	χ^2	卡方临界值		GI
				0.975	0.025	
A	随机	0.78	7.8	2.7	19.0	0.00
B	集聚	6.07	54.6		>19.0	0.17
C	均匀	0.07	0.6	<2.7		-0.03

表 3.2 三种分布指数成分,在最大程度均匀性、随机性和最大程度的集聚性每个指数的值,
摘自 Elliott(1973)^a

指数	指数值在		
	最大均匀性	随机性	最大集聚性
分布指数, s^2/\bar{x}	0	1	n
集聚指数, $s^2/\bar{x} - 1$	-1	0	n-1
格林指数, $(s^2/\bar{x}) - 1/(N-1)$	-1/(N-1)	0	1

a: \bar{x} = 每个 SU 个体平均数, s^2 = 方差, n = 样本中个体总数

正如前面讨论所见到的那样,对于估测一组数据与泊阿松分布的一致性,方差与平均数比率(ID)是一个十分有用的统计检验。但在表示集聚程度时, ID 并不是十分有用的。在表 3.2 我们显示了在最大均匀性(即每个取样单位包含相同数目的个体)、随机性和最大程度的集聚性时(即所有个体被发现在单一的 SU) ID 值。当种群是集聚的时候, ID 强烈地受这个样本中个体数目的影响(即 n, 表 3.2), 所以仅仅在每个样本中 n 相同时, ID 才是一个有用的, 用来比较集聚程度的指数(Elliott 1973)。

指数 2: 集聚指数。 David 和 More(1954)提出了 ID 的修改形式并命名为“集聚指数”(IC), IC 由下式计算:

$$IC = (s^2/\bar{x}) - 1 = ID - 1 \quad (3.9)$$

对于随机的格局 IC 等于 0, 当有最大的规则性时为 -1, 当有最大的集聚性时为 n-1(表 3.2)。与 ID 的情形类似, 由于它对于 n 的依赖性(表 3.2) 所以作为集聚程度的比较测定并不是一个十分有用的集聚指数。这个评论对于 ID 的许多变体是适合的(Elliott 1973)。下一个指数是这种情况的一个例外。

指数 3: 格林指数。格林(1966)提出了一个对 IC 的修改形式。它独立于 n, 由下式计算:

$$GI = \frac{(s^2/\bar{x}) - 1}{n - 1} = \frac{IC}{n - 1} \quad (3.10)$$

GI 在 0(随机)和 1(最大集聚)之间变化(表 3.2), 这样格林指数可以被用来比较个体总数、样本的平均数和样本中 SU 的数目都不同的样本。因此, 在提出的测定集聚程度众多 ID 的变体中, GI 看来是最值得推荐的。表 3.1 给出了图 3.1 中介壳虫种群的 GI 值。

3.3 例子:木蜂幼虫

在南新墨西哥高丝兰的花序上木蜂幼虫是常见的。一位对这些蜜蜂幼虫感兴趣的昆虫生态学家在 180 枝高丝兰茎上收集了蜜蜂幼虫的样本。这些数据以频度表的形式总结如下:

x:	0	1	2	3	4	5	6	7	8	9	10
F _x :	114	25	15	10	6	5	2	1	1	0	1

x 为每个花序柄上蜜蜂幼虫的数目, F_x 是含有 x=0,1,2,...,r 个幼虫的高丝兰花序柄的频度,在这个例子中 r=10,总的 SUs 数目的是:

$$N = \sum_{x=0}^{10} (F_x) = 114 + 25 + \dots + 0 + 1 = 180$$

个体总数是:

$$n = \sum_{x=0}^{10} (x F_x) = (0)(114) + (1)(25) + \dots + (9)(0) + (10)(1) = 171$$

样本的算术平均数是:

$$\bar{x} = n/N = 171/180 = 0.95$$

方差是:

$$s^2 = [\sum_{x=0}^{10} (x F_x)^2 - \bar{x} n]/(n - 1) = [681 - (0.95)(171)]/179 = 2.897$$

3.3.1 泊阿松分布

步骤 1. 假设。 这个零假设就是每枝花序柄上木蜂幼虫的数目遵从泊阿松分布,因此是随机的格局,在预先对这些蜜蜂空间格局的不甚了解时,这个假设看来是有道理的。

步骤 2. 频度分布, F_x。 观测的频度分布、平均数和方差如前面给出的那样。

步骤 3. 泊阿松概率, P(x) 表 3.3a 给出了每枝花序柄上发现 x=0,1,2 和 3 个木蜂个体的以及累积的概率。这些值是在公式(3.1)用平均数 0.95 获得。注意,在一个 SU 中发现 3 个个体后的累积概率是 98.4%,所以,剩余的概率 P(4)到 P(10)将仅为 1.6%,通过这些不同性,这些频度级就被划分了。

$$P(4^+) = 1.0 - 0.984 = 0.016$$

表 3.3a 每个 SU 具有 x 个体的泊阿松概率 P(x)的计算(高丝兰花序柄)

		累积概率
P(0)	=e ^{-0.95}	=0.3867
P(1)	=(0.95/1)P(0)	=(0.9500)(0.3867)=0.3674
P(2)	=(0.95/2)P(1)	=(0.4750)(0.3674)=0.1745
P(3)	=(0.95/3)P(2)	=(0.3167)(0.1745)=0.0553
P(4 ⁺)	=1.0-0.9839	=0.0161

表 3. 3b 含有 x 个蜜蜂 SUs 期望频度 E_x 的计算

				累积 SUs
E_0	$= (N)P(0)$	$= (180)(0.387) =$	69.61	69.61
E_1	$= (N)P(1)$	$= (180)(0.367) =$	66.13	135.74
E_2	$= (N)P(2)$	$= (180)(0.174) =$	31.41	167.15
E_3	$= (N)P(3)$	$= (180)(0.055) =$	9.95	177.10
E_{4+}	$= (N)P(4+)$	$= (180)(0.017) =$	2.90	180.00

表 3. 3c 卡方检验统计量 χ^2 的计算

每花序柄上幼虫数	频度		
	(x)	F_x	E_x
	0	114	69.6
	1	25	66.1
	2	15	31.4
	3	10	9.9
	4+	16	2.9
总和		180	180.0
			$\chi^2 = 121.8$

步骤 4. 期望频度, E_x 。 期望频度是由泊阿松概率与 SUs 总数相乘获得(表 3. 3b), 分配后的频度级数为 $q=5$ 。

步骤 5. 吻合性。 卡方检验统计量(χ^2)计算如下(见表 3. 3c)。

$$\begin{aligned}\chi^2 &= [(114 - 69.6)2/69.6] + \dots + [(16 - 2.9)2/2.9] \\ &= 28.3 + \dots + 59.3 = 121.8\end{aligned}$$

将这个统计量与自由度为 $q-2=3$ 的卡方概率表相比较, 获得 χ^2 值高达 121.8 的概率小于 1%, 所以我们拒绝这个零假设, 即每个花序柄上蜜蜂幼虫的数目遵从泊阿松分布, 也就是随机分布。注意在这个例子中最小的频度期望值为 2.9, 按着 3. 1. 1 节的讨论(步骤 3)这个值同我们曾允许的最小值(即 1)相比要大一些, 但同我们推荐的当 $q=5$ 时最小值为 3 相比要小。变化允许的最小期望值对 χ^2 检验统计量有什么影响呢? BASIC 程序 POISSON.BAS(软盘中)可以对这一问题进行快速的检验, 因此我们推荐学生按着程序中的说明对这一问题作进一步的分析。

3. 3. 2 负二项分布

步骤 1. 假设。 这个零假设就是高丝兰花序柄上的蜜蜂幼虫呈集聚格局, 因此就可检

验(每个SU 个体的数目)与负二项分布的一致性。由于方差大于平均数,所以有理由怀疑是集聚格局。

步骤 2. 频度分布, F_x 。(见 3.3.1 节)

步骤 3. 负二项概率, $P(x)$ 。利用公式(3.5)和 $\bar{x}=0.95, s^2=2.897$ 获得 \hat{k} 的估计值是:

$$\hat{k} = (0.95)^2 / (2.897 - 0.95) = 0.4635$$

由于 \hat{k} 和 \bar{x} 小于 1, 公式(3.4)应被用来估计 \hat{k} 。将 $N=180$ 和 $N_0=114$ 代入公式(3.4)左侧(LHS)得值为 0.1984, 再将 $\hat{k}=0.4635$ 代入方程(3.4)右侧(RHS)得出下面结果:

$$\hat{k} \log_{10}(1 + \bar{x}/\hat{k}) = 0.4635 \log_{10}(1 + 0.95/0.4635)$$

$$= 0.2245$$

由于 RHS 值大于 0.1984, 将一个小于 0.4635 的 k 值代入公式(3.4), 选择 $\hat{k}=0.30$, 给出下列结果:

迭代 2:

$$\hat{k} \log_{10}(1 + \bar{x}/\hat{k}) = 0.30 \log_{10}(1 + 0.95/0.30)$$

$$= 0.1859$$

这个值与 0.1984 值较为接近(但低一点), 所以对下一个迭代, 选择一个较高的 \hat{k} 值, 使用 $\hat{k}=0.34$ 得

迭代 3:

$$\hat{k} \log_{10}(1 + \bar{x}/\hat{k}) = 0.34 \log_{10}(1 + 0.95/0.34)$$

$$= 0.1969$$

再一次, 对下一个迭代, 我们将试一个较高的 \hat{k} 值, $\hat{k}=0.3457$

迭代 4:

$$\hat{k} \log_{10}(1 + \bar{x}/\hat{k}) = 0.3457 \log_{10}(1 + 0.95/0.3457)$$

$$= 0.1984$$

这个数字与公式(3.4)的 LHS 恒等。因此对于这个例子, \hat{k} 的最佳估计值是 0.3457, 在 BASIC 程序 NEGBINOM 使用与上面步骤相同的逐步方案来估测 \hat{k} 。

接下来, 利用公式(3.3), 在表 3.4a 中给出了每个花序柄上 $[\bar{x}=0.95, \hat{k}=0.3457, \bar{x}/(\bar{x}+\hat{k})=0.7332]$ 发现有 0, 1, 2, 3 幼虫个体的和累积概率, 在一个 SU 发现 4 个个体后的累积概率为 94.6%, 所以剩余的概率, $P(5)$ 到 $P(10)$ 将仅占 5.4%, 也就是 $P(5^+) = 1.0 - 0.946 = 0.054$, 正如将在步骤 4 中看到的那样, 这个级别的划分保证了期望值没有落在 5.0 下面的值里。

步骤 4. 期望频度, E_x 。某一级别的期望频度是由这一级别的概率与 SUs 总和相乘获得。

步骤 5. 吻合性。卡方检验统计量(χ^2)计算如下(见表 3.4c)

$$\chi^2 = [(114 - 114.00)^2 / 114.00] + \dots + [(10 - 9.67)^2 / 9.67] \\ = 0.00 + \dots + 0.01 = 1.18$$

表 3. 4a 每个取样单位含有 x 个个体的负二项分布概率 P(x) 的计算

		概率	累积概率
P(0)	$= [1 + (0.95/0.3457)]^{-0.3457}$	= 0.6333	0.6333
P(1)	$= (0.7332)(0.3457/1)P(0) = (0.2535)(0.6333)$	= 0.1605	0.7938
P(2)	$= (0.7332)(1.3457/2)P(1) = (0.4933)(0.1605)$	= 0.0792	0.8730
P(3)	$= (0.7332)(2.3457/3)P(2) = (0.5733)(0.0792)$	= 0.0454	0.9184
P(4)	$= (0.7332)(3.3457/4)P(3) = (0.6133)(0.0454)$	= 0.0278	0.9462
P(5 ⁺)	$= 1.00 - 0.9462$	= 0.0538	1.0000

表 3. 4b 含有 x 个体 SU_s 的期望频度 E_x 计算

				累积 SU
E ₀	$= (N)P(0)$	$= (180)(0.633) =$	114.00	114.00
E ₁	$= (N)P(1)$	$= (180)(0.161) =$	28.90	142.90
E ₂	$= (N)P(2)$	$= (180)(0.079) =$	14.25	157.20
E ₃	$= (N)P(3)$	$= (180)(0.045) =$	8.17	165.30
E ₄	$= (N)P(4)$	$= (180)(0.028) =$	5.00	170.30
E ₅₊	$= (N)P(5^+)$	$= (180)(0.054) =$	9.68	180.00

表 3. 4c 卡方统计量的计算, χ^2

每枝花序柄	频度		
幼虫数目 (x)	F _x	E _x	$(F_x - E_x)^2 / E_x$
0	114	114.0	0.00
1	25	28.9	0.53
2	15	14.3	0.04
3	10	8.2	0.41
4	6	5.0	0.19
5 ⁺	10	9.7	0.01
总计	180	180.0	$\chi^2 = 1.18$

将这个值与自由度为 q-3 即 3 卡方分布表中的临界值相比较,在 5% 概率水平上的卡方临界值是 7.82。由于获得卡方值为 1.18 的概率远远低于这个值,所以我们不能拒绝零假设。看来负二项型比较适合这组观测数据,在作出木蜂幼虫事实上呈集聚格局的确切阐述之前,我们不想作进一步的肯定(例如一组独立的数据)。注意在这个例子中当最小的期望允许低到 1 和 3 时, χ^2 值分别是 2.6 和 2.5, 还是低于这个临界值。

3.3.3 分布指数

木蜂数据的分布指数是[来自公式(3.6)]:

$$ID = 2.897/0.95 = 3.05$$

由于 $N > 30$, d 是一个确切的检验统计量, 计算如下(公式 3.8):

$$d = \sqrt{2 \times 545.86} - \sqrt{2 \times (180 - 1) - 1} = 33.04 - 18.89 = 14.15$$

由于 $d > 1.96$, 与泊阿松分布的一致性被拒绝, 并建议为集聚分布, 它与前面描述的吻合性检验相一致(3.3.1 节和 3.3.2 节)。

集聚程度可以通过计算格林指数获得[公式(3.0)]:

$$GI = (3.05 - 1)/(171 - 1) = 0.012$$

由于 GI 的最大值为 1.0(即所有 171 个个体出现在一单一的高丝兰花序柄上), 故这个值代表一个相对低的集聚程度。

3.4 例子: 杨树上的织蛾

一位森林病理学家研究了寄生于威斯康星森林上的织蛾, 获得了杨树随机 100 个枝条样本的下列数据:

$x:$	0	1	2	3	4	5	6
$F_x:$	24	29	26	74	5	0	2

表 3.5 每个枝条上织蛾的观测频度分布与按着泊阿松和负二项分布($\hat{k}=8.75$)概率给出的期望频度一致性的卡方吻合性检验, 这些结果是使用程序 POISSON.BAS 和 NEGBINOM.BAS 获得的

统计量和指数总结						
SU 数目						100
个体数目						155
每个 SU 个体数目: 平均数	=1.55					
方差	=1.73					
分布指数(方差: 平均数):	1.11					
χ^2 统计量[公式(3.7)]:	110.2					
d 统计量[公式(3.8)]:	0.81					
格林指数:	0.00073					
		泊阿松		负二项		
织蛾数/枝	观测枝条数	期望	χ^2	期望	χ^2	
0	24	21.2	0.36	24.0	0.00	
1	29	32.9	0.46	31.6	0.21	
2	26	25.5	0.01	23.2	0.34	
3	14	13.2	0.05	12.5	0.18	
4+	7	7.2	0.01	8.7	0.33	
	100	100.0	$\chi^2=0.89$	100.0	$\chi^2=1.06$	
			(df=3)			(df=2)

x 是每个枝条上织蛾的数目, F_x 是含有 x 个织蛾的枝条数目。为说明 BASIC 程序 POIS-

x 是每片叶子上红螨的数目, F_x 是具有 x 个螨叶子的频度。利用 BASIC 程序 POISSON. BAS 和 NEGBINOM. BAS 即泊阿松和负二项分布模型用来配合这组数据以作为例证(表 3.6), 使用吻合性检验 ($P < 0.001$), $\chi^2 = 26.6$, $df = 2$) 和 d 统计量 ($|d| = 7.1$ 它大于 1.96) 均明显地拒绝了与泊阿松模型的一致性, 接近于 2.0 的 ID 值表明它是一个集聚格局, 而且事实上负二项模型对这组数据确实配合的很密切 ($\chi^2 = 2.45$, 它小于 $df = 3$, $P < 0.05$ 的临界卡方表值), 按照格林指数测定 ($GI = 0.0057$) 集聚的程度比较低。

3.6 分布模型的附加讨论

种群空间格局的一个成分就是它的强度(Intensity)。Pielou(1977)定义强度为: 出现在一个样本中不同密度的范围。例如在图 3.1 中植物 B 上昆虫格局强度(集聚分布)同植物 A(随机分布)的相比要高, 一个值得考虑的有趣问题就是平均密度对测定格局强度各种指数的影响, 也就是随机变疏对我们所提供的各种指数值的影响是什么。例如在图 3.1 中通过随机地移走每个植物上个体的一些成分来变化平均密度, 很明显我们不想使格局的指数对在密度方面的变化太敏感, 关于这种类型的问题, Myers(1978)指出区别统计产物与明显的生物学效果的重要性。

Lloyd(1967)发展了他所称的聚块指数(Index of patchiness), 它是用来测定格局的强度而不受随机变疏的影响。首先平均拥挤 \bar{x} (Mean crowding) 定义为在一个 SU 中每种个体的其它个体的平均数目, 由下式计算:

$$\bar{x} = \bar{x} + IC \tag{3.11}$$

IC 是集聚指数

Lloyd 定义聚块程度为

$$\bar{x}/\bar{x} = (\bar{x} + IC)/\bar{x} \tag{3.12}$$

Hill(1973a)表示凡是类似这种形式的任何指数:

$$(\text{方差} - \text{平均数})/\text{平均数}^2$$

将大致不受随机变疏的影响。Morisita(1971)也提出了一个指数 Morisita 指数, I_d 。这个指数也不受密度随机变疏的影响。Morisita 指数与 Lloyd 的聚块指数几乎是平行的, 也就是

$$I_d = \left(\frac{n}{n-1}\right)\left(\frac{\bar{x}}{\bar{x}}\right) \tag{3.13}$$

同以前一样, n 是在这个样本中的个体的总数。

Myers(1978)发现格林指数(GI)[公式(3.10)]与平均密度没有关系, 并推荐它用于分析随着密度的变化有机体空间格局的变化上。我们对 GI 的检验表明在种群是集聚的时候, 支持 Myers 的推荐。但当种群空间格局是随机的或均匀的时候, GI 看来是与平均密度有关的。

负二项分布参数 \hat{k} 在随机变疏减少种群密度的情况下是另一个基本保持不变的参数。但应该强调的是只有当亲本种群适合负二项分布时, \hat{k} 参数的这个性质才是真实的。Pielou(1977)指出 \hat{k} 也可以写成分布指数(ID)或集聚指数(IC)的形式:

$$\hat{k} = \frac{\bar{x}}{ID - 1} = \frac{\bar{x}}{IC}$$

\hat{k} 的倒数(即 $1/\hat{k}$)也是常常作为一个指数使用的,它也是(方差-平均数)/平均数² 的形式。在最大程度集聚情况下, \hat{k} 趋向于 0,但 \hat{k} 并不是个理想的指数,这是由于它受 SUs 总数和样本中个体数目的影响(Southwood 1978),在一个样本中,随着样本容量或样本中个体总数的增加, \hat{k} 趋向于 0。

3.7 总结和推荐

1. 本章所提供的格局检验方法使用的基本数据是由每个 SU 的个体组成的,SUs 被认为是自然出现的生境单位(3.1 节)。

2. 每个 SU 的个体数目与方差之间的关系受有机体分布潜在格局的影响,泊阿松分布(平均数=方差)被推荐为随机分布模型,负二项分布为集聚分布模型(3.1 节)。

3. 当样本的容量(SUs 数目) >30 ,数据(含有 $0, 1, 2, \dots, r$ 个个体取样单位的数目)应该安排为频度分布形式。对这些数据配合的泊阿松和负二项分布模型的成功与否应经过卡方吻合性检验(3.2.1 节和 3.2.2 节的步骤 5)和 d 统计量的计算。应记住的一点是拒绝随机的和非随机的假设失败后,也许只意味着使用我们手头上的特定数据检验一些其它的格局的存在失败了。如有可能,独立的数据应被应用以确定你的发现。

4. 只有关于空间格局站得住脚的假设才能被提出,而且你的结论也不应单单建立在统计显著性检验的基础上(3.1 节),回忆一下,同样的数据可以用不同的模型来描述(见 3.4 节中的例子)。

5. 当样本的容量小于 30,分布指数(方差与平均数比)被推荐去检验数据与理论泊阿松模型的一致性(3.2.3 节)。

6. 各种指数,建立在每个取样单位个体数目的平均数和方差比被提出测定一个种群的集聚程度,只有格林指数,它是独立于在一个样本中个体总数的,被推荐用于不同群落指数的比较(除非每个群落个体的计数是相同的,这种情况稀少)。

第四章 样方方差法

在前面一章里,我们描述了一些在不连续生境中种群空间格局的统计模型的使用,这些不连续的生境为自然的取样单位(SUs),如腐烂的园木、植物的叶子。在这一章里我们将涉及那些连续地通过植物群落的种类个体的空间格局(如森林中的树木)。由于生境是连续的或者说是非间断的,要获得一个样本就必须使用一些人为的取样单位,因而所得到的结果将会受到选择的SU形状和大小的影响。这个问题的提出使得一些方法得到发展,并且使我们能够检验不同大小的SU对于所定的潜在空间格局的影响。总的来说,这些被称为样方方差方法。

4.1 基本方法

在所研究的地域内,当一个种类个体的分布是连续的(如贯穿于一群落中的一种草),获得一个样本必须使用人为的SU。举个例子,考虑一下如图4.1所描述的那样,在一个群落内一个样带或邻接格子样方被放置和观测,注意如个体的空间格局是随机的,每个样方个体的观测值与集聚的和均匀的格局相比将相当不同(当然,平均数和方差也一样)。本章所描述的方法是以这种连续的样方的样本获得的数据为基础的,并且利用这些数据提出群落中与个体的空间格局有关的假设。

每当人为的SUs(样方)在取样过程中被利用,认识到样方的大小和形状对结果的影响是十分必要的。对于图4.1b中所显示的集聚分布,很明显每个样方中个体的数目将受到加倍样方尺寸的严重影响,另一方面,当分布是随机的时候(图4.1a),加倍样方的尺寸并不是什么问题,每个样方中个体的期望数目(不考虑大小,只要所有样方大小一样)贯穿一随机种群的将是相同的,每个样方中个体数目的频度分布将总是遵循普阿松分布(Pielou 1977)。样方方差方法是以检验通过一系列不同的SU尺寸的每个SU个体数目的平均数和方差的变化为基础的。

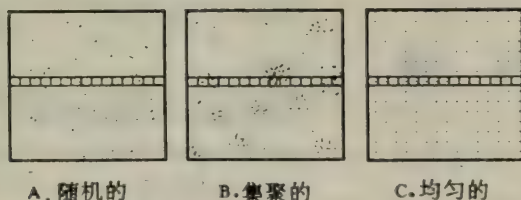


图 4.1 在某一群落于取样地区个体是(a)随机分布(b)集聚或(c)均匀分布连续样方(一条样带)的放置

本书将提供两类样方方差分析方法,其一以区组样方方差(Blocked-quadrat variance, BQV)为基础,另一类以成对样方方差(Paired-quadrat variance PQV)为基础。BQV的方法主要是采用 Greig-Smith(1983)和他的学生 Kershaw(1973), Hill(1973a)和 Usher(1975)的研究工作, PQV 方法的过程是来自 Goodall(1974), Ludwig 和 Goodall(1978)的研究工作。BQV 方

法是通过利用区组或结合相邻或邻接的样方尺寸的变化来确定格局的强度(密度的变化范围)和格局的纹理(个体斑块间的距离),PQV 方法是利用样方间隔的变化提供这种信息。

4.2 步骤

对于下面所描述空间格局的分析方法,让我们假定对一连续的群落采用一条连续样方的样带(如图 4.1 所表示的那样)取样。注意邻接的格子使用基本相同的方法将研究地域分成 N 个相等的样方。对于放在这个群落的连续的样方带,可以观测每个感兴趣的种在每个样方中个体的数目(或盖度,生物量等等),这条样带数据可以表示为一观测的向量 $X=x_1, x_2, x_3, \dots, x_N$, N 是观测样方的总数。

4.2.1 区组样方方差法

使用这个方法,我们计算在不同“区组规模(面积)”上个体数目的方差。不同的区组面积是通过前面所描述的方式对 N 个样方逐步地结合获得的。例如,图 4.2 给出了 8 个连续样方样带的区组顺序,在区组规模 1 中,有 8 个连续样方,在区组面积 2 中,邻接的样方合并成 4 个长方形的样方,它的大小为区组规模 1 的 2 倍,通过合并区组规模 2 的邻接样方形成下一个区组

连续样方的样带							
1	2	3	4	5	6	7	8
区组样方方法		区组规模	样方				
BQV		1	(1)(2)(3)(4)(5)(6)(7)(8)				
		2	(1,2)(3,4)(5,6)(7,8)				
		4	(1,2,3,4)(5,6,7,8)				
TTLQV		1	(1)(2)(3)(4)(5)(6)(7)(8)				
		2	(1,2)(3,4)(5,6)(7,8)				
		3	(1,2,3)(4,5,6)				
		4	(1,2,3,4)(5,6,7,8)				
成对样方方法		间隔	样方				
PQV		1	(1)(2)(3)(4)(5)(6)(7)(8)				
		2	(1,2)(2,3)(3,4)(4,5)(5,6)(6,7)(7,8)				
		3	(1,3)(2,4)(3,5)(4,6)(5,7)(6,8)				
		4	(1,4)(2,5)(3,6)(4,7)(5,8)				
		5	(1,5)(2,6)(3,7)(4,8)				
		6	(1,6)(2,7)(3,8)				
		7	(1,7)(2,8)				

图 4.2 8 个连续(邻接)样方组成的样带。给出了使用在不同区组规模或间隔对于 BQV,TTLQV 和 PQV 方法的方差计算的样方分配方案
规模 4。由于一对对邻接的样方被合并,故 BQV 区组规模以 2 的幂的形式增长(即 $2^1=2, 2^2=4, 2^3=8, 2^4=16, 2^5=32, 2^6=64, 2^7=128, 2^8=256$ 等等):如图 4.2 显示的那样,另一个区组方

案是可能的(以后描述的 TTLQV 方法),以及成对方案(4.2.2 节描述的 PQV 方法)

步骤 1. 在区组规模 1 由下式计算 BQV 方法方差:

$$VAR(X)1 = (2/N) \{ [1/2(x_1 - x_2)^2] + [1/2(x_3 - x_4)^2] + \dots + [1/2(x_{N-1} - x_N)^2] \} \quad (4.1a)$$

步骤 2. 数据是从相邻成对的样方组合为长方形的两样方区组获得的,它是 2 倍于原来的一样方区组大小,区组规模 2 的方差由下式计算:

$$VAR(X)2 = (4/N) \{ [1/4(x_1 + x_2 - x_3 - x_4)^2] + [1/4(x_5 + x_6 - x_7 - x_8)^2] + \dots + [1/4(x_{N-3} + x_{N-2} - x_{N-1} - x_N)^2] \} \quad (4.1b)$$

步骤 3. 在较高区组规模(如 8, 16, 32, 64 等)方差的计算可参照方程(4.1)(同时参见 Ludwig 和 Goodall 1978),当然最大可能的区组规模由于所研究的样带大小(或长度)确定(图 4.2)。

步骤 4. 绘制获得的方差与其各自对应的区组规模图。

正如我们刚刚阐述的那样,由 Goodall(1954a)和 Greig-Smith(1954a)最初发展的 BQV 方法仅测定 2 的幂数的方差。这样,使用在样带上有用的样方数目必须限制在 2 的某些幂的范围内(如 $2^8=256$)。为了避免这一限制,Hill(1973a)发展了另一种区组方案,这种方案就是两项局部样方方差[Two-term local quadrat variance(TTLQV)]方法,这种 TTLQV 方法应该代替前面所描述的没作修改的 BQV 方法的使用。

TTLQV 在区组规模 1 的方差计算是

$$VAR(X)1 = [1/(N-1)] \{ [1/2(x_1 - x_2)^2] + [1/2(x_2 - x_3)^2] + \dots + [1/2(x_{N-1} - x_N)^2] \} \quad (4.2a)$$

区组规模 2 由下式计算

$$VAR(X)2 = [1/(N-3)] \{ [1/4(x_1 + x_2 - x_3 - x_4)^2] + [1/4(x_2 + x_3 - x_4 - x_5)^2] + \dots + [1/4(x_{N-3} + x_{N-2} - x_{N-1} - x_N)^2] \} \quad (4.2b)$$

在逐次变大的区组(图 4.2,同时参见 Ludwig 和 Goodall 1976)的方差计算类似公式(4.2)。基本来说,Hill 的 TTLQV 方法估测方差对于所有的区组规模适合选择的区组大小为 $N/2$ (推荐 $N/10$)。不管怎样,很明显 TTLQV 方法是 BQV 的变体,因为样方沿着样带仍然是区组的。

4.2.2 成对样方方差法

在 PQV 方法中,沿样带以特定间隔或距离选择成对样方估测方差(图 4.2)。由于样方大小是固定的,样方间的间隔或距离是产生方差的仅有成份,而不象 BQV 的情形,样方的间隔

和大小看上去都是产生方差的因素。事实上样方的大小和间隔对方差的影响很可能不同,这就告诉我们分开这两种影响要比混淆它们的影响更好一些(Goodall 1974)

步骤 1. 在间隔 1 按下式计算 PQV 方法方差

$$VAR(X)1 = [1/(N-1)]\{[1/2(x_1 - x_2)^2] + [1/2(x_2 - x_3)^2] + \dots + [1/2(x_{N-1} - x_N)^2]\} \quad (4.3a)$$

注意:这个计算与 TTLQV 在区组规模 1 的计算是相同的[图 4.2, 方程(4.2a)]

步骤 2. 间隔 2 的方差由下式给出:

$$VAR(X)2 = [1/(N-2)]\{[1/2(x_1 - x_3)^2] + 1/2[(x_2 - x_4)^2] + \dots + [1/2(x_{N-2} - x_N)^2]\} \quad (4.3b)$$

步骤 3. 对于较大间隔方差的计算类似公式(4.3), 适宜的间隔应小于最大间隔 $N-1$ (推荐 N 的 10%)。

步骤 4 绘制由 PQV 方法获得的方差与对应的间隔图。

在 PQV 方法中, 计算了在给定间隔所有可能成对的样方方差, 另一个步骤是利用随机成对样方(RPQV)估测方差(Goodall 1974), 这种样方对是以给定的间隔从所有可能成对的样方以随机的方式选择的。随机选取是没有替换的, 所以一旦成对样方被选择, 它们就不能在这个间隔或其它任何选择的间隔中被利用。这个过程的优点在于每个间隔方差的估测是独立的, 而且可以计算统计概率以检验在这些不同空间间隔方差不同的显著性。与之对比, 以一个间隔利用所有样方对计算方差忽略了以不同间隔不同方差估测的独立性, 因此任何这种统计检验都是无效的。

4.2.3 方差图的解释

利用 TTLQV 或 PQV 方法获得的方差和与之对应的区组规模和间隔图将表现出三种格局之一(图 4.3), 或着某些可能的中间类型趋势。

1. 随机的。 在所研究的地域内, 如果研究的个体的分布是随机的, 方差将随着不同的区组规模或间隔随机地变动(图 4.3a)。

2. 均匀的。 如果这些个体是均匀分布的, 方差将很低而且在不同的区组规模或间隔的方差将趋向于不变。

3. 集聚的。 如果所研究的个体是集聚的, 在与集聚平均面积(由每个聚块所占据的平均面积)的半径等同处的区组规模或间隔处, 方差将趋向于高峰, 聚块中心与相邻聚块中心间的平均距离将是这个区组规模或间隔的 2 倍。如果这些聚块它们自己是规则格局化的, 方差的峰值将随着在这个峰值之外的区组规模或间隔的增加而降低为一个较低的值(图 4.3c)。如果方差的峰值高而陡, 这表明这个集聚的格局具有较高的强度和较强的纹理。换句话说, 格局存在着一个紧密的明显的聚块, 聚块之间存在着较大的开阔空间, 如果通过一些区组规模或间隔出现的峰值较低, 那么这个格局的强度就较低, 而且聚块也不能很好地被确定, 给出的例子将阐述其它的集聚格局。

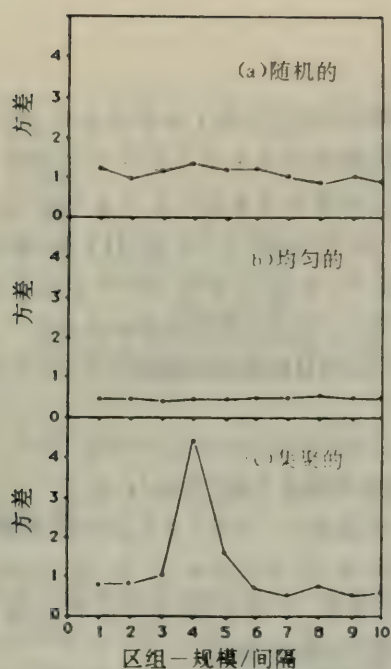


图 4.3 与区组规模/间隔相对应的(a)随机的,(b)均匀的,(c)集聚空间格局方差的典型图

4.3 举例:计算

利用一模拟的单一数字的向量,我们阐述一下 SPA 的两类样方方差的计算过程,在一个 $N=8$ 连续样方的样带个体数目为:

$$X = (x_1, x_2, x_3, \dots, x_8) = (1, 0, 3, 5, 2, 1, 4, 6)$$

现在我们叙述所推荐的区组样方 SPA 方法 Hill 的 TTLQV 技术,接着阐述所推荐的成对一样方 SPA 方法的 Ludwig 和 Goodall 的 PQV 技术。

4.3.1 TTLQV 方法

步骤 1. 在区组规模 1,测定区组样方方差[公式(4.2a)]

$$\begin{aligned} \text{VAR}(X)_1 &= [1/(8-1)]\{[1/2(1-0)^2 + [1/2(0-3)^2] + \dots + [1/2(4-6)^2]\} \\ &= 1/7[(0.5)(1) + (0.5)(9) + \dots + (0.5)(4)] \\ &= 1/7(0.5 + 4.5 + \dots + 2.0) = 1 \div 7 \times 18.5 = 2.64 \end{aligned}$$

步骤 2. 区组规模 2 的方差(方程 4.2b)

$$\begin{aligned} \text{VAR}(X)_2 &= [1/(8-3)]\{[1/4(1+0-3-5)^2] + [1/4(0+3-5-2)^2] \\ &\quad + \dots + [1/4(2+1-4-6)^2]\} \\ &= 1/5[(0.25)(49) + (0.25)(16) + \dots + (0.25)(49)] \\ &= 1/5(12.25 + 4.0 + \dots + 12.25) = 1 \div 5 \times 35.75 \\ &= 7.15 \end{aligned}$$

步骤 3. 区组规模 3 和 4 的方差

$$\begin{aligned}
 VAR(X)3 &= [1/(8-5)]\{[1/6(1+0+3-5-2-1)^2] \\
 &\quad + [1/6(0+3+5-2-1-4)^2] \\
 &\quad + [1/6(3+5+2-1-4-6)^2]\} \\
 &= 1/3[(0.17)(16) + (0.17)(1) + (0.17)(1)] \\
 &= 1/3(2.67 + 0.17 + 0.17) = 1 \div 3 \times 3.00 = 1.00 \\
 VAR(X)4 &= [1/(8-7)]\{[1/8(1+0+3+5-2-1-4-6)^2]\} \\
 &= (1/1)[1/8(16)] = (1/1)(2.0) = 2.00
 \end{aligned}$$

4.3.2 PQV 方法

步骤 1. 在间隔 1, 测定成对样方方差[方程(4.3a)]:

$$\begin{aligned}
 VAR(X)1 &= [1/(8-1)]\{[1/2(1-0)^2] + [1/2(0-3)^2] + \dots + [1/2(4-6)^2]\} \\
 &= 1/7[(0.5)(1) + (0.5)(9) + \dots + (0.5)(4)] \\
 &= 1/7(0.5 + 4.5 + \dots + 2.0) = 1 \div 7 \times 18.5 = 2.64
 \end{aligned}$$

步骤 2. 在间隔 2 的方差[方程(4.3b)]:

$$\begin{aligned}
 VAR(X) &= [1/(8-2)]\{[1/2(1-3)^2] + [1/2(0-5)^2] + \dots + [1/2(1-6)^2]\} \\
 &= 1/6[(0.5)(4) + (0.5)(25) + \dots + (0.5)(25)] \\
 &= 1/6(2.0 + 12.5 + \dots + 12.5) = 1 \div 6 \times 37.5 = 6.25
 \end{aligned}$$

步骤 3. 在间隔 3 和 4 的方差:

$$\begin{aligned}
 VAR(X)3 &= [1/(8-3)]\{[1/2(1-5)^2] + [1/2(0-2)^2] + \dots + [1/2(2-6)^2]\} \\
 &= 1/5[(0.5)(16) + (0.5)(4) + \dots + (0.5)(16)] \\
 &= 1/5(8.0 + 2.0 + \dots + 8.0) = 1 \div 20.5 = 4.10
 \end{aligned}$$

$$\begin{aligned}
 VAR(X)4 &= [1/(8-4)]\{[1/2(1-2)^2] + [1/2(0-1)^2] + \dots + [1/2(5-6)^2]\} \\
 &= 1/4[(0.5)(1) + (0.5)(1) + \dots + (0.5)(1)] \\
 &= 1/4(0.5 + 0.5 + \dots + 0.5) = 1 \div 4 \times 2.0 = 0.5
 \end{aligned}$$

使用 TTLQV 方法, 本例中可以获得方差估测值的最大区组规模是 4 (由于 $N=8$, 故最大区组规模值为 $N/2$, 如图 4.2 所示)。对于这个例子我们也阐述了方差计算的最适间隔为 4, 尽管可以测定最高可达 $N-1$ 间隔的方差 (如图 4.2 所表示的那样, 本例中 $N=8$, 最大间隔为 7), 但由于随着区组规模和间隔的增加, 自由度降低, 在较大的区组规模和间隔上对方差的估测缺乏可靠性, 所以我们建议在取得规模或间隔超过 N 的 10% 时方差就不能再计算了 (例如, 如果 $N=500$, 在区组规模和间隔超过 50 的, 就不能再计算方差了)。这个基本规则是相对保守的, 增加到 N 的 20% 可能不会很大地降低精确性。

步骤 4. 对应间隔绘制方差图 (图 4.4)。

4.3.3 方差图的解释

在步骤 4, 使用 TTLQV 和 PQV 方法获得的方差对应于区组规模或间隔绘制成图 (图 4.4)。这个简单的例子表明, 这个聚集斑块大小是以 2 个样方为半径, 根据是方差的峰值在区组规模 2 和间隔 2, 这样聚块之间是以 4 个样方宽为间隔 (平均), 也就是, 与峰值有关的 2 倍于样方区组规模或间隔 (峰值在 2 的 2 倍处 = 4)。

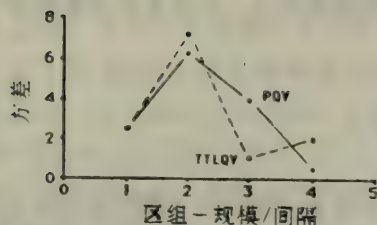


图 4.4 本例计算中对应于区组规模/间隔的方差图

4.4 例子:荒漠灌丛中的麻雀巢

下一个例子,我们举的是一位生态学家在 Chihuahuan 荒漠灌丛地对黑颈麻雀巢空间格局的研究。取样采用一条 10m 宽,2600m 长的样带,每 10m 为一单位,记录每一单位内巢的数目,总计给出 260 个取样单位(SUs),沿着这条样带,每个 SU 巢的观测数目的向量是:

```
X=(0 1 5 2 1 0 0 1 0 0 2 5 1 0 0 0 1 0 0 1 0 2 6 1 0 1 0 0 0 0
  1 3 2 1 0 0 0 0 0 0 1 0 0 2 4 1 0 0 1 0 0 1 0 0 0 3 5 7 2 1
  1 0 0 0 1 0 0 0 3 4 1 0 0 0 4 3 1 0 0 0 0 0 1 0 0 0 0 3 6 2
  2 1 0 0 0 0 0 4 1 0 0 0 0 1 0 0 0 0 0 1 0 1 4 2 1 0 0 0 3 7
  4 2 0 1 0 0 0 1 0 0 0 0 1 1 0 1 5 2 0 0 1 0 0 1 1 0 1 1 4 5
  4 2 1 0 0 0 0 0 0 1 1 5 4 1 0 0 0 1 0 0 1 0 1 0 1 4 3 2 0 0
  0 1 0 0 0 0 1 0 0 1 3 2 0 0 1 2 0 0 1 0 0 0 0 0 3 6 2 1 0 0
  0 0 1 0 4 6 3 1 0 0 0 1 0 1 0 0 0 3 5 4 0 0 1 0 1 0 0 0 2 3
  5 2 1 0 0 1 0 0 0 2 0 0 0 3 6 2 0 2 0 0)
```

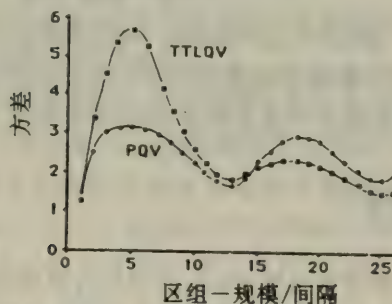


图 4.5 在荒漠灌丛地使用区组样方方差法(TTLQV)和成对样方方差法对鸟巢分布的空间格局分析

利用软盘中的 BASIC 程序 BQV.BAS 和 PQV.BAS 计算以不同区组规模使用 TTLQV 和以不同间隔使用 PQV 方法的方差,方差与其对应的区组规模和间隔(适宜数为 26, SUs N 的 10%)见图 4.5。

当在聚块之间与聚块内部这一段空间里鸟巢密度有个较大的差异时,在区组规模/间隔 5

处出现了一个方差峰值,也就是,当对比聚块之间较低密度或空的样方有几个高密度的样方落在一聚块内,最大的方差峰值也就出现了。这鸟巢的集聚就有一个较强的指示,即 100 米间隔有个聚块。

4.5 例子:草场上的火蚁

最后一个例子,我们举一个昆虫生态学家在佛罗里达草场上以一条 10 米宽,1280 长的样带对火蚁丘空间格局研究的例子(取样以 10m 为一单位,总计给出 128 个 SUs)。在每个 SU 观测到的蚁丘数目的观测向量是

$X=(1\ 2\ 0\ 3\ 4\ 6\ 4\ 0\ 1\ 0\ 1\ 0\ 3\ 6\ 7\ 5\ 2\ 0\ 1\ 0\ 0\ 3\ 7\ 4\ 2\ 0\ 1\ 1$
 $0\ 2\ 4\ 6\ 5\ 2\ 1\ 0\ 0\ 1\ 2\ 4\ 7\ 3\ 0\ 0\ 1\ 0\ 1\ 0\ 3\ 6\ 5\ 2\ 0\ 0\ 0\ 1$
 $0\ 1\ 3\ 8\ 4\ 1\ 0\ 0\ 1\ 1\ 0\ 5\ 6\ 3\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 4\ 7\ 4\ 2\ 0\ 0\ 1$
 $0\ 0\ 1\ 0\ 5\ 7\ 2\ 1\ 0\ 0\ 0\ 1\ 0\ 4\ 6\ 3\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 2\ 5\ 8\ 3\ 2$
 $0\ 1\ 2\ 4\ 7\ 5\ 3\ 1\ 0\ 0\ 0\ 2\ 1\ 3\ 8\ 4)$

利用 BASIC 程序 BQV.BAS 和 PQV.BAS 计算 TTLQV 和 PQV,与区组规模或间隔(适合值为 13,也就是 128 的 10%)相对应的方差图表示在图 4.6 中

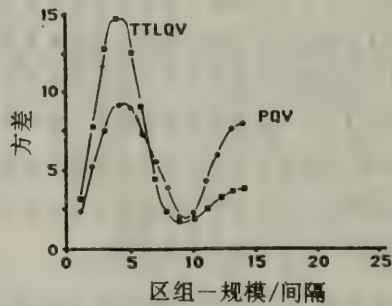


图 4.6 使用 TTLQV 和 PQV 对蚁丘空间格局分析

区组样方和成对样方方差图都强烈地表明火蚁丘在此草场上呈集聚分布(注意 PQV 峰值在 4.5 处),蚁丘聚块之间的距离或间隙大约为 90m(即 2 倍于 4.5 倍的 10m)。

4.6 样方方差法的附加讨论

样方方差法方面的发展是针对于克服 Goodall(1954a)和 Greig-Smith(1952a)在 BQV 方法上的不足,这个早期方法的最大不足是以连续的 2 的幂去估测区组的规模(也就是 1,2,4,8,16,32,64 等等),Hill(1973a)的 TTLQV 方法有效地克服了这个不足,所以,我们建议用它来代替原来的 BQV 方法,这个 TTLQV 方法准确地测定了格局的强度和规模(Ludwig 和 Goodall 1978)它甚至可以检测出两个格局规模或水平的集聚,尽管相对于较小规模的格局它更强调较大规模的格局(Ludwig 1979)。

BQV 方法的另一个不足是区组的初始位置可以影响结果,特别是当有一个规则格局存在

时,Usher(1975)使用一组在两个规模上(4和30)带有明显的格局的人为数据,应用BQV方法和他对BQV的修改形式称之为逐步区组样方方差(Stepped blocked-quadrat Variance, SBQV)方法,逐步的过程减少了这样一些情况即事实上格局为规则的时候初始位置的影响(Usher 1975, 1983)。Usher的在4和30两个规模格局的人为数据也用RPQV、PQV和TTLQV方法分析,Ludwig和Goodall(1978图1)讨论和显示了比较结果。这里简单地重复一下主要的发现,TTLQV、PQV和RPQV方法在较低程度上显示规模4的格局在30的格局内呈振荡状态;使用TTLQV方法小规模格局在大规模格局内是平滑的,这是因为方差的计算是由区组包括几个重复的小规模格局计算的;在PQV方法中,小规模(格局)的振荡是无阻尼的(这是因为每个方差仅是由一对样方产生),而且聚块也是固定大小的,RPQV方法在处理大数目的间隙时由于其精确性较低,所以这种方法也只是大致遵循PQV方法获得的期望;RPQV方法应该用在检验其它方法获得的方差峰值结果的显著性上。

Carpenter和Chaney(1983)应用模拟的格局数据也比较了SPA不同的样方方差法,他们发现Goodall的RPQV方法(以8个重复随机成对为基础)得出的方差与其对应的区组规模或间隔图正确地给出了平均聚块大小,比TTLQV、BQV或光谱分析要准确。在他们的比较中,没有使用PQV方法,但PQV(以所有可能成对的样方为基础)同RPQV(以没有替换的随机成对样方为基础)相比将会给出更准确的方差估计值。他们发现BQV和TTLQV方法混淆了斑块和间隙的长度,因而过分地估计了斑块的大小。

在一个整体的格局中,如果有人想知道一个特定的方差峰值是否显著地(统计上的)高于邻近的波谷,那么就需要独立地估测方差。所有的区组样方方差法(BQV, SBQV和TTLQV)和PQV方法都不能提供在不同区组规模或间隔上独立的方差估测,这是因为在每个不同区组或间隔上每个方差的计算都是由这完整的一组数据来计算的。

估测独立方差的一种方法是应用Goodall的RPQV方法,这种方法估测方差是通过从样方的样本中随机地选取一对样方(没有替换)而进行的,但我们也应当注意这种方法不能给出真正的统计的独立估测,通过没有替换的取样,每当一个样方从剩余的样方组中剔除,一个新的条件概率就产生了(Zahl 1977)。但随着取样规模的增加(比如说超过200),方差的估计将趋向于独立,当然在极端条件下的集聚,应该强调的是没有替换随机取样的影响,因为移走一个高强度样方的机会将会严重地影响条件概率。

Ludwig和Goodall(1978)建议随机地将连续样方数据分成两个亚组,其中一亚组用PQV方法探索其格局,另一亚组用RPQV方法来检验在任何方差峰值(在选定的间隔对峰值和它的两侧)的显著性(我们支持这个推荐为SPA的合理过程;我们还得重复前面提到的注意事项,即在小样本的规模上用RPQV方法进行的方差估测也许不是这样)。这个随机分组过程可以进行较大数目的重复以获得对给定方差的“误差统计”。这种“蒙特-卡罗”式的过程被Carpenter和Chaney(1983)成功地运用到计算随机成对的样方方差。他们同时也建议使用来自独立的、随机的格局(零模型)的期望方差作为测定峰值和谷值显著性的一个参照点。

除了方差以外,分布的测定也可以在SPA中应用,如第三章描述的Morisita(1959)指数, I_d ,由于它相对独立于平均密度,这样在一个集聚个体的种群中,如果计算了 I_d ,并将它对其所对应的区组规模绘图,直到区组规模与平均聚块尺寸相等前将保持相对恒定,之后增加。

4.7 总结和推荐

1. 当你对于一个出现在连续地段种群的空间格局是否是非随机的(均匀的或集聚的)检验感兴趣时,一个人为 SUs 的样本结合“样方方差”方法可以被利用,通过样方大小和间隔的变化可以提供有关格局的信息,Kershaw(1973),Pielou(1974,1977)和 Greig-Smith(1983)都对群落生态学中样方(人为的 SUs)取样的应用进行过讨论。

2. 我们建议 Hill 的 TTLQV 方法可以应用于连续样方的数据,因为它能够准确地检验格局的规模和强度,PQV 方法也可以应用于同样的数据,因为它同 TTLQV 区组过程相象,没有混淆样方大小和间隔。TTLQV 和 PQV 分析的结果应该是互补的,两个结果间的任何不同都可能提供额外的格局信息,如前面提到的两个规模格局的出现(Ludwig 和 Goodall 1978,Ludwig 1979)。

3. 不能用超过 N (样方总数)的 10% 的区组或间隔来计算方差(对 TTLQV, PQV 均适合)。

4. 利用“方差齐性”假设,使用 Goodall 的 RPQV 方法在选出间隔小的组上可以检验峰值(表示集聚)的统计显著性,这里方差不同的显著性是在峰值和峰谷之间。

第五章 距离方法

在前面一章里,我们提供了由连续样方获得的取样数据,也就是用一系列邻接的通过生物群落(如森林或草地)放置的小样方观测值进行空间格局(SPA)的分析方法。对于一个必须使用人为 SUs 的群落,可以替换上述方法的就是应用无样地(Plotless)或聚离方法(Distance method)。在这一章里我们将描述怎样用获得的距离数据测定种群的空间格局,从点与点,个体与个体之间的距离可以导出一些指数,而且可以应用检验统计量来检验潜在空间格局的假设。

5.1 基本方法

距离或无样地取样技术是由植物生态学家在对植被类型获得迅速估测的过程中发展起来的,在这些植被类型中,植物个体能够被很好地定义并能明显地区别开来,如在一片森林中(Cottam 和 Curtis 1956)。在个体是稀少而广泛地分散情况下,距离取样同样方相比显得更有效,如在一个大的样方中寻找和计数个体是相当费时间的(当个体稀少时,应避免使用空样方),Cottam 和 Curtis 描述了一些基本的取样过程,如点中心四分法和最近邻体法。

一些其它的距离取样过程连同由距离数据导出的格局的各种指数已被介绍(见 Diggle 1983)。在众多的可能方法中,T 形取样过程通过比较模拟格局和绘制野外数据图提供了一个强有力的格局指数(Diggle 等 1976, Diggle 1983, Lamacroft 1983)。在这一章里,我们将提供 T 形取样技术并阐述它是怎样为空间格局测定和随机分布的显著性检验统计量提供数据的。象将要被描述的那样,位于一个群落内部每个取样点 T 形取样技术需要测定两个距离:(1)从这个点到最近个体间的距离,(2)从那个个体到与其最近的个体间的距离。

在这一章里,我们也提供了另一个分布指数(见 3.2.2 节),它是以点到个体间的距离为基础的,这个系数被证明它在测定空间格局方面是强有力的方法(Johnson 和 Zimmer 1925)。我们建议大家回顾一下 3.3.2 节,以强调以自然 SUs(第三章)和以距离(这章里)为基础的空间格局测定的不同。

Diggle 指出,仅以点到个体的距离为基础的所有格局指数和统计量是值得“怀疑”的,因为它们不能区别单一个体和点丛(个体非常紧密的聚块)。因为随机的或网格内的点趋向于不落在个体紧密的聚块内,点到个体间的距离通常测定的是点到那些聚块边缘个体的距离,这样,点丛边缘个体同单一个体将不易区别,我们赞成 Diggle 等(1976)指出的那样,即使样方一方差方法(第四章)提供较多的 SPA 信息(即可以同时检验几个规模的格局),但距离方法对于最初 SPA 的研究仍不失为一种特别有价值的技术(即当某个群落的格局需要较快和较容易的大致结果时)。

5.2 步骤

5.2.1 T 形距离取样

在一个群落随机地或在一规则网格内选取取样点,在每个随机点上测定两个距离,如图 5.1 所示,首先测定从点(O)到最近个体(P)(任何方向)的距离 X ,然后在 P 点处作一条垂直 OP 的直线(T 形的顶部,图 5.1 的虚线),个体(P)到它的最近邻体(Q)的距离 Y 要在由这个垂线产生的“半个飞机”的另一侧产生。

如果通过一连续地段种群个体的格局是完全随机的,那么点到个体距离的平方期望值将近似等于最近邻体距离平方的一半,即: $E(x^2) = 1/2E(y^2)$ 。但如果个体的格局是集聚的时候,点到个体距离平方的期望值将大于最近邻体距离平方的一半,即 $E(x^2) > 1/2E(y^2)$ 。均匀格局的情况与此正好相反即 $E(x^2) < 1/2E(y^2)$ [注意:考虑平方的距离是自然的,因为距离的测定是通过二维空间的(面积 $= \pi x^2$; (图 5.1))。如果这些平方的距离表示为比率的形式,这种比率就可以被用作空间格局指数[如公式(5.1)]和作为一个检验统计量,正如我们将要阐述的那样。

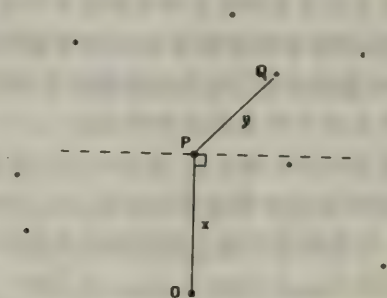


图 5.1 T 形取样距离: x , 点到与它最邻近个体间 P 的距离。 y , 个体 P 到与它最邻近个体 Q 的距离, 选择 Q 时要避免与 O 在“半个飞机”(虚线以下)同侧

5.2.3 空间格局的 T 形指数

从获得的 T 形取样距离可以导出一个空间格局指数(C),它是一个点到个体距离 X 的平方与个体到最近邻体距离 Y 的平方的比率,即

$$C = \frac{\sum_{i=1}^N [X_i^2 / (X_i^2 + 1/2 Y_i^2)]}{N} \quad (5.1)$$

N 是取样点的总数,对于随机的格局 C 接近 $1/2$,均匀格局显著小于 $1/2$,集聚格局显著大于 $1/2$,对于 C 值偏离 $1/2$ 的显著性检验,可用 z 值计算

$$z = \frac{C - 0.5}{\sqrt{1/(12N)}} \quad (5.2)$$

由于用 $1/(12N)$ 作为方差估测时 C 是接近于正态分布,因此 C 的统计显著性可以由标准正态分布概率表(在 $P=0.05, z=1.96$)确定 z 的显著性获得。

5.2.3 分布的距离指数(I)

Eberhardt(1967)和 Pielou(1959,1977)以及其他一些人描述了各种分布指数,这些指数仅需要点到最邻近个体间的距离,但这些指数都有各种各样的局限性(Goodall 和 West 1979),例如 Pielou 指数需要对种群密度独立的估计(不是单一的东西),Goodall 和 West 都表示他们的两个指数(以及其它的)同以距离对(即点到个体和个体到最邻近个体的距离)为基础的指数相比,能力不够。

Johnson 和 Zimmer(1985)提出一个新的分布距离指数,它也仅以点到个体的距离为基础,这个指数看来对空间格局的检验也还是比较有效的。对于给定 N 个点的样本和从点 i 到与其最近的个体间的距离 x_i ,分布的距离指数(I)可以定义为平方距离的平方总和与平方距离总和的平方的比率,方程如下,它包含了许多平方。

$$I = (N + 1) \frac{\sum_{i=1}^N (x_i^2)^2}{[\sum_{i=1}^N (x_i^2)]^2} \quad (5.3)$$

Johnson 和 Zimmer 指出对于随机的格局来说,I 的期望值近似于 2,小于 2 为均匀格局,大于 2 为集聚格局,他们进一步指出在中等取样规模时(当相当接近于 100 时)I 收敛于正态,它提供了 z 值的基础。

$$z = \frac{I - 2}{\sqrt{4(N - 1)/(N + 2)(N + 3)}} \quad (5.4)$$

这个 z 要与标准正态分布临界值相比较,以获得背离随机格局的显著性。

5.3 举例:计算

我们采用一组简单的,人为的 10 个点到个体(x_i)和个体到邻体(y_i)距离对(以米表示)来阐述这章提供的两个距离方法的计算。

点	1	2	3	4	5	6	7	8	9	10
x_i	6	2	6	5	6	2	1	2	4	2
y_i	4	9	3	7	1	8	3	5	6	5

5.3.1 空间格局的 T 形指数(C)

步骤 1. 使用公式(5.1)计算 C:

$$C = \left[\frac{6^2}{6^2 + 4^2/2} + \frac{2^2}{2^2 + 9^2/2} + \dots + \frac{2^2}{2^2 + 5^2/2} \right] / 10$$

$$= (0.82 + 0.09 + \dots + 0.24) / 10 = 0.454$$

步骤 2. 使用公式(5.2)计算 C 的显著性

$$z = \frac{0.454 - 0.5}{\sqrt{1/[(12)(10)]}} = \frac{-0.046}{0.091} = -0.51$$

对于这个 C 值,其检验值 z 小于 z 的临界值 1.96,我们接受这个假设,即所测定的点到个体间的距离是在一个潜在的随机格局之内。

5.3.2 分布距离指数(I)

步骤 1. I 的计算需要点到个体距离平方的平方总和:

$$\begin{aligned}\sum_{i=1}^{10} (x_i^2)^2 &= (6^2)^2 + (2^2)^2 + \dots + (2^2)^2 = 36^2 + 4^2 + \dots + 4^2 \\ &= 1296 + 16 + \dots + 16 = 4834\end{aligned}$$

步骤 2. 计算点到个体距离平方总和的平方

$$\begin{aligned}\sum_{i=1}^{10} (x_i^2) &= 6^2 + 2^2 + \dots + 2^2 = 36 + 4 + \dots + 4 = 166 \\ \left[\sum_{i=1}^{10} (x_i^2) \right]^2 &= 166^2 = 27556\end{aligned}$$

步骤 3. 利用公式(5.3)计算 I

$$I = [(10 + 1)(4834)]/27556 = 1.93$$

这个指数接近于随机格局的期望值 z

步骤 4. I 的显著性,应用方程(5.4)计算检验统计量 z

$$\begin{aligned}z &= (1.93 - 2) / \sqrt{[4(10 - 1)]/[(10 + 2)(10 + 3)]} \\ &= -0.07 / \sqrt{36/(12)(13)} = -0.07 / \sqrt{0.23} = -0.15\end{aligned}$$

将这个 z 值同标准正态分布表中的临界值相比较获得背离随机的显著水平,仅仅偶然获得 $z = -0.15$ 的概率就非常大于 $P = 0.05$,因此随机格局的假设是可以接受的。

5.4 例子:松林

下一个例子,我们举的是一个森林生态学家利用无样地取样对北卡罗来纳中麓火炬松林研究的例子,获得下列点到树(x_i)和 T 形最近邻体(树到树)距离(y_i):

树 _i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	6.1	5.2	3.6	1.5	1.6	2.2	4.1	3.2	3.1	1.2	3.5	3.6
y_i	6.4	6.1	6.3	6.3	6.1	5.8	6.3	6.5	6.1	5.5	5.6	6.4
树 _i	13	14	15	16	17	18	19	20	21	22	23	
x_i	2.1	6.5	3.5	2.1	3.3	1.4	6.5	4.6	3.3	6.1	3.6	
y_i	5.6	5.5	6.4	6.3	5.6	6.4	6.3	5.5	6.3	5.5	6.2	

用 BASIC 程序 TSQUARE. BAS(软盘中)计算空间格局 T 形指数 C , 分布距离指数 I , 统计量 z , 结果是 $C=0.387, z=-1.877; I=1.767, z=-0.634$, 这个集聚指数小于 $1/2$, 分布指数小于 2, 表明松树有趋向于均匀格局的趋势, 对于 C 的 z 值(忽略符号)小于 1.96 表明在这种情况下没有明显背离随机性而趋向于均匀性。对应于 I 的 z 值, 获得它的概率也大于 5%, 所以建议我们接受这个松树林是随机格局的零假设。

使用模拟的格局对 I 和 C 能力的比较研究中, Ludwig 和 Diggle 发现对于均匀格局 I 显得不足, 而 C 却强有力, 这些没有发表的发现支持 Diggle 等(1976)的结果, 即 C 在检验均匀性的格局时是强有力的, 因此在均匀性的情况下, 我们将接受 C 的结果。

5.5 荒漠中的收获蚁

最后一个例子, 让我们举一个荒漠群落中收获蚁群体空间格局测定的例子, 测定了 20 个随机点到最近群体距离(x_i)和与 i 相关的群体到群体间的距离(y_i), 以米表示的距离数据如下:

Pt _i	1	2	3	4	5	6	7	8	9	10
x_i	6.5	7.4	9.2	4.8	5.2	6.0	5.6	9.2	0.4	6.5
y_i	1.4	2.9	0.9	1.7	2.5	2.3	3.4	8.1	1.1	2.9
Pt _i	11	12	13	14	15	16	17	18	19	20
x_i	7.4	7.6	3.3	4.6	7.2	9.9	8.3	8.0	6.9	7.5
y_i	0.3	6.2	0.6	2.5	3.6	9.1	1.8	2.6	0.9	1.9

再一次利用 BASIC 程序 TSQUARE. BAS 并获得以下结果: $C=0.872, z=5.761$ 和 $I=1.327, z=-1.737$, 在随机格局这个零假设的条件下获得 C 值为 0.872 的概率小于 1%, 所以拒绝这个零假设, 由于 $C>1/2$ 表明格局集聚强烈, 但与之相对比, 分布的距离指数 I 小于 2, 由它得出的结论为格局是均匀的, 这是怎么回事? 注意一下, 绝大多数点到个体间的距离(x)在 4—10m 范围内, 而绝大多数个体到个体间的距离(y)小于 4m, 这表明随机点落在了点丛(小

的,紧密聚块)之间, x 是从这些点到点丛边缘距离,但 y 很短,是点丛内部个体间距离。换句话说,收获蚁群体有“多元出入”(点丛),这个分布距离指数 I 表示这些点丛(群体)是均匀的格局,尽管这个例子表明 I 在区分单一个体和点丛时是无能为力的,但它确实表明 I 值提供了额外的空间格局信息,本例中也就是点丛它们自己也许是均匀的格局。

5.6 距离方法的附加讨论

使用已知空间格局的人为种群,Goodall 和 West(1979)比较了检验格局的 8 种指数的能力,在所有非随机格局检验中他们发现 Hopking 指数能力超过 80%(成功检验),并得这样的结论,在所有的距离指数中,Hopking 聚集指数是最有能力的,但在他们的比较中并不包括 T 形指数。

Hopking 指数的能力也被其它的研究工作如 Diggle(1983)所肯定,他曾指出“就能力方面讲,Hopking 指数树立了一个标准,别的方法不及它”,既然如此,你也许会问,这一章里为什么不提供(或推荐)Hopking 指数呢?答案很简单,就是 Hopking 指数的效力的检验依赖于个体到最邻近个体中的个体是“真正”的随机选择,这就使得 Hopking 指数显得不实用,由于个体随机取样要求将所有个体计数和标记以便随机选择(Diggle 1983),这就使问题变得很明显,总数的计数正是生态学家所想避免的。正如 Pielou(1977)所强调的那样,使用随机点去选择随机个体,并不是合理的过程,因为一个强烈集聚的种群随机点的使用将会极大地偏向于选择聚块与聚块中间的个体。

以距离(点到个体,个体到邻体距离)对为基础的另一个格局指数是条件距离比率方法(Cox 和 Lewis 1976),这种方法使用的距离是从取样点到最近个体距离(x)和从这些个体到最近邻体距离(y),但没有象 T 形取样那样限制在“半个飞机”的另一侧,这些距离被“条件为”(分成)两组。第一组就是邻体的距离小于 2 倍的点到个体间的距离($y_i < 2x_i$),第二组就是邻体的距离大于 2 倍的点到个体的距离($y_i > 2x_i$)。Cox 和 Lewis 指出,几乎关于格局的所有信息存在于第一组数据中,并导出带有参数的指数用以描述集聚的格局,Lamacroft(1983)基本肯定了条件距离方法(对于模拟和绘制在图上的种群密度估计),但也发现 T 形方法更为健全而且偏见少。5.4 节中曾提到的 Ludwig 和 Diggle 的比较研究中发现,T 形距离和条件距离指数在检验集聚的和均匀的格局时是有效的,所以本章提供了计算较简单的 T 形指数,Ludwig 和 Diggle 同时还发现,非随机地简单地选择取样点并不影响这些指数的能力,例如以规则的或相等间隔网格点选取取样点。

象 Johnson 和 Zimmer 分布距离指数那样,Eberhardt(1967)指数也是以取样点到最近个体的距离为基础,但它在格局为集聚的情况下表现欠佳。Goodall 和 West(1979)发现 Eberhardt 指数在检验小规模集聚时成功率较低,大约为 31%左右,而对于规则格局检验的成功率可达 96%。

Pielou(1959)也提出一个指数,它仅需要点到个体的距离,但需要估测种群密度。这个密度的估测被规定为独立于距离的测定(如使用样方取样)。Johnson 和 Zimmer(1985)发现,对于几个检验模型他们的分布指数 I (它是独立于密度的)同 Pielou 的相比更为有效。

Clark 和 Evans 指数(1954)也需要独立的密度估测,它是以最近邻体的距离为基础,而不

是点到个体间的距离,Clark 和 Evan 指数经常偏向于规则分布,但这个偏差可以被纠正(Sinclair 1985)。Goodall 和 West(1979)并没有肯定 Holgate(1965)的最近邻体指数,发现这个指数同其它指数相比效果差,特别是同 Hopking 指数相比。

5.7 总结和推荐

1. 距离方法提供了对出现在连续群落中个体空间格局快速测定的方法。尤其是当那些个体易于辨认的时候(森林中的树),距离可以从规则的网格选取,即可不是随机的,这将不影响出现在本章中系数的效果。

2. 推荐了以 T 形取样法获得的距离数据比为基础的空间格局指数。

3. 推荐了由 Johnson 和 Zimmer(1985)描述的分布距离指数,它仅测定点到最邻近个体间的距离,正如 5.5 节中阐述的例子那样,仅以点到个体距离为基础的格局指数不能区别单一的分散的个体和在紧密聚块边缘的个体。

4. 我们建议对空间格局的 T 形指数 C 和分布距离系数 I 都要进行计算,5.5 节中阐述了 C 是怎样提供了关于小的紧密的集聚信息和 I 是怎样提供关于这些聚块的空间结构的其它信息的。

第三部分 种—多度关系

第六章 背景

在生态群落中观察的最显著和连续的现象之一就是种—多度的变化。这个变化导致生态学家提出与群落本质有关的中心问题。例如在一给定群落中有多少种类,它的相对多度为多少,多少种是稀少的,多少种是常见的,种的相对多度格局的数学描述产生了有关群落稳定性、能流、资源配置、种—面积关系和进化过程的理论(如 Hutchinson 1959, Kolasa 和 Bicsiadka 1984, McGuinness 1984, McNaughton 和 Wolf 1970, Whittaker 1965, 1972)。

种多度通常以每个种个体数为基础,但其它一些变量如生物量、盖度也可以被利用。事实上它们也许更重要,这些多度数据被处理成种多度分布形式,也就是包含 $x=1, 2, 3, \dots, r$ 个个体种的数目的频度分布(Pielou 1975),为配合这些种的多度数据,一些统计模型得到了发展(见 Whittaker 1972),在仅需少的、易估测的和生态学上可以解释的参数就可找出一个“普遍模型”的希望促使下,这项工作得到了发展。通过比较在不同群落中获得配合多度的参数估测值,生态学家将有一个研究群落本质的工具(Hendrickson 1979),在第七章,一个这样的模型即对数正态将被描述。

第八章,介绍多样性指数,这些指数提供了在一个群落中描述物种多度关系的变化的方法。存在于生态学文献中的多样性看来有两个含义:(1)一个群落中种的总数 r (经常称为种的丰富度)和(2)多样性的双重概念,它是种的丰富度和相对多度的结合。

Whittaker(1972)定义了生态学家感兴趣的多样性的三个特殊水平(1) α 或生境内的多样性;(2) β 或生境间的多样性(即随着环境梯度的变化);(3) γ 或大规模景观多样性(α 和 β 多样性的结合),第八章将讨论测定 α 多样性的问题。

6.1 矩阵展示

	取样单位										群落取样
	1	2	3	4	5	N	
种	a										
	b										
	c										
	.										
	s										
环境因子	w										
	x										
	.										
	z										

图 6.1 阴影部分表示物种多度关系的生态数据形式,在此感兴趣的是在单一 SU 或取样单位平均值(群落样本)的物种多度

表 3.1 给出了测定物种多度关系的生态矩阵形式,重点放在列上,一个大范围内的数据组是可能的。例如,种多度的频度分布(第七章)可能是来自对于有机体集合的单一观测(如用一个扫网在大面积的草场上捕获的昆虫)。或者也许是一些扫网[取样单位(SU)]被使用,并且利用这些组数据获得频度分布。在使用多样性指数方面(第八章),单一 SU(如上面的那个扫网收集物)的多样性可以被计算,或者一些 SUs 的平均值也可以被计算以代表一个群落的样本,与一个群落样本多样性相对的单一 SU 的多样性问题我们将在第八章中讨论。

6.2 选择的文献

应用种多度模型的一些选择的生态学文献的例子在表 6.1 中已给出,多样性指数和分布模型被应用在从水生到陆地群落的广泛数据中。

表 6.1 种多度关系模型选择使用的文献例子,DU = 分布模型;DI = 多样性指数

地点	群落	方法	参考文献
西澳大利亚	展欧石楠	DU	Lamont 等 1977
亚得里亚海	浮游植物	DI	Revelante 和 Gilmartin 1980
南英格兰	泰晤士海湾	DU	Andrews 和 Richard 1980
柯林赤海(美国)	水生昆虫	DI	Kaesler 和 Herricks 1976
南卡罗莱纳	两栖动物	DI	Bennett 等 1980
马来亚	热带森林	DU	Kempton 和 Taylor 1978
大不列颠	林地	DI	Claridge 和 Wilson 1981
死海,以色列	一年生植物	DI	Danin 1976
英格兰	海底生物	DI	Mason 1977
佛罗里达	落叶林	DI	Monk 1967
	针叶林	DI	Del Moral 1977
北卡罗莱纳	落叶林	DI	Christensen 1977
西班牙	草地	DI	de Pablo 等 1982
三门河	水生无脊椎动物	DU	Minshall 等 1985
东北部(美国)	林鸟	DU	Able 和 Noon 1976

第七章 分布模型

当从一个生态群落中获得了动物或植物种多度大的样本时,可以以各种方式总结这些数据,以帮助检验多度和具有某个多度种的数目的关系,我们可以使用代表有 $1, 2, 3, \dots, r$ 个体种的数目的频度分布(或其它别的多度方法,如生物量、盖度等)。另一个方法是绘制这个群落中按着种序列由最大到最小的种多度图。当以这种方式总结种多度数据时,一定的格局就出现了。在想要获得种多度关系的经验描述和在检验生态群落有机体潜在组织的假设这两种愿望的促进下,生态学家提出了各种模型以试图使这些多度格局数量化。

7.1 基本方法

在一个群落中,当每个种的相对多度按着由最丰盛的到最不丰盛的种序列排列起来绘制成图时,我们可以观察到几种特征分布。由于群落是由大量种集合组成的,故结果总象图 7.1 曲线 b 所描述的那样(May 1975, 1981)。曲线 7.1b 显示的数据来自 Whittaker(1965)在亚利桑那州,圣卡他利那山种类丰富的荒漠群落的研究工作。在那里,众多的种具有中等程度的多

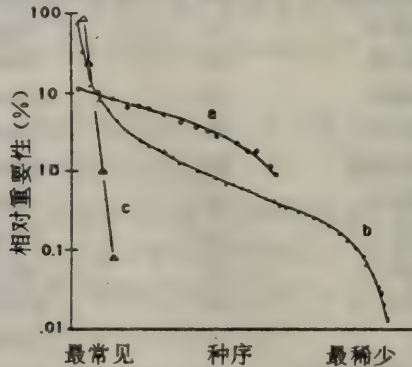


图 7.1 群落中种的相对重要性(以比例为基础的测定)和它们的种序列(由最常见到最稀少)的关系,曲线 a 是落叶林中雏鸟对密度的断棍分布,曲线 b 是荒漠群落植物地表盖度的对数正态分布,曲线 c 是高海拔冷杉林多度的几何分布

度(以地面盖度形式测定),只有少数几个种具有较低或较高的多度,人们从各种各样的生态群落中包括硅藻、土壤节肢动物、哺乳动物群、鸟和植物获得了有机体与上面类似的空间格局(见 Sugihara 1988 的回顾)

同曲线 7.1b 相比,两个其它的多度格局被认为是极端情况,它们是从种类相对少、均匀的群落样本中获得的。第一种极端情况如 7.1 曲线 c 所描述的那样,它同曲线 7.1b 相比,其多度的分布是一个明显的不平分布,这种格局的一个例子就是 Whittaker(1965)在烟山高海拔的冷杉林的研究。他发现这些冷杉林的特性就是种类非常少,总多度(由净第一性生产力测定)的绝大多数是由数目非常少的种类所贡献的,第二种极端情况如图 7.1 曲线 a 所描述的那样,它比曲线 7.1b 明显平,这最后的一种格局描述的是西佛吉尼亚落叶林雏鸟对密度,在此群落中总

的多度由不同种类相对平均地分配,自然中可以观察到这种情况的仅有几例(Giller 1984, Whittaker 1970)

随着这些观察而来的一些明显问题是:在与生态群落组织有关的种多度格局方面是怎样不同的(McNaughton 和 Wolf 1970),我们这里所提到的群落“组织”是围绕以生态位(Ecological niche)概念为中心的一些观点。Whittaker(1972)定义生态位概念为群落中种的“位置”,包括资源利用、活动时间、水平和垂直的位置、种群间的作用方式等等。因此生态学家发展了几个假设以试图用生态位的形式解释这些观测到的多度关系。图 7.1 显示了由这些假设而产生的三种理论分布:几何分布(曲线 c),对数正态分布(曲线 b),和“断棍”分布(曲线 a)(尽管在这一点我们没有提出,读者应当认识到这些分布的每一种都有相应的数学描述)。简而言之,按着图 7.1 显示出的不同分布,提出了三种群落“组织”过程。

1. 几何分布(Geometric distribution). 想象在一个物种相对贫乏的群落中,单一的环境资源(如湿度)对于物种的生存是极端重要的,并以十分严格的等级制度利用着,一个单一的优势种优先占领这种资源的大部分,紧接下来的种优先占领剩余资源的一小部分等等,这种要求导致了种多度的单一几何分布(曲线 7.1c),这个模型也称“生态位优先占领假说”。

2. 断棍分布(Broken-stick distribution). 如果我们假设在一个群落中,湿度资源是一条棍,如果我们随机地将这条棍折成数段并分别为这个群落的每一个种利用,那么我们就得到了种丰盛度所谓的断棍分布(曲线 7.1a)。这个模型假定在一个群落中种间分布或利用一些重要的资源是没有重叠的。

3. 对数正态分布(Lognormal distribution). 给定一大量种集合,它们的相对多度很可能是许多或多或少独立的因子相互作用的产物(May 1981),因此,这样的种多度关系可由对数正态分布定量化的描述,对数正态分布是统计的中心极限定理的一个结果(May 1975)。正如前面所提到的那样,这种对数正态分布在大多数群落多度格局的描述时得到应用。

本章提供对数分布的一个介绍,而不考虑几何或断棍模型这两种特殊情况,有关群落组织和生态位理论的详细介绍,包括一些难懂东西的阅读我们推荐 Whittaker(1970),May(1975)和 Giller(1984)。

最初,发展这些模型背后的一些动力就是寻找一个物种多度关系的一个普遍模型,在这样的模型中可以通过参数的不同性和相似性很容易地比较各种群落(Giller 1984)。给定的这个模型事实上可以代表与生态群落潜在格局有关的正式假设(正如前面我们所接触到的),以获得关于种的生态位和种是怎样共存或分享可利用的环境资源的基本信息。对生态学家来说,一个普遍模型将是一个有用的工具,但看来没有这样的一个基本范例存在。对提出的各种多度模型的仔细研究可以发现:矛盾的假设可以导致产生同样的模型,进一步讲,从抵触是假设中导出的不同模型用来配合同样的观测数据(Giller 1984, May 1981, Poole 1974)(回忆一下我们在第三章注意到的与之相类似的问题,对作为潜在空间格局统计分布的解释),因此,我们获得的这些模型是检验格局或趋势的工具,它们对于产生关于群落可以检验的假设也许是有用的(May 1975)。

7.2 步骤

为了提供对数正态模型,我们首先对图 7.1 中曲线 b 的数据重新安排,即以种的频度(纵坐标)对应多度级的对数(横坐标)绘图。结果如图 7.2。Preston(1948,1962)介绍了在生态学中对数正态分布的使用,利用以 2 为底的对数,这样每一级(或倍频程)就是前一个多度级 2 倍,这是人为的,但现在在生态学研究这是一个方便的步骤。

图 7.2 的数据看上去是遵循对数正态分布的。正如我们早些时候注意到的那样,May (1975)表示这是不足为奇的,因为大的,不同种类的集合的相对多度趋向于由许多独立的因子控制,根据统计的中心极限定理,分布将是对数正态的。May 表示事实上,任何大的,不同类的集合其对数正态分布是真实的,同时他还用了美国财富和世界各国人口分布的例子。生态群落中种的多度关系,通常是遵循对数正态分布的,这已被公认(Sugihara 1980)

其对数正态分布由下式给出

$$S(R) = S_0 e^{-a^2 R^2} \quad (7.1)$$

$S(R)$ 是由众数起第 R 个倍频程种的数目, S_0 是在众数倍频程上种数目的估计值(该倍频程上种数最多),参数 a 是一个分布宽度的倒数测量(即 $a=1/2\sigma$, σ 是标准差)。

步骤 1. 观察的频度分布。将观测的数据处理为频度分布的形式,给出了在每个多度级或倍频程上种的数目,依照 Preston 以 2 为底表示多度的惯例,第 i 个倍频程上 R 值由下式给出。

$$R = \log_2(N_i/N_0) \quad (7.2)$$

N_i 是在第 i 个倍频程上种的多度, N_0 是众数倍频程上种多度,由公式 7.2 可以看出在众数倍频程上 R 总是 0,由于 $N_i=N_0$ 而 $R=\log_2(1)$,所以 $R=0$ 。

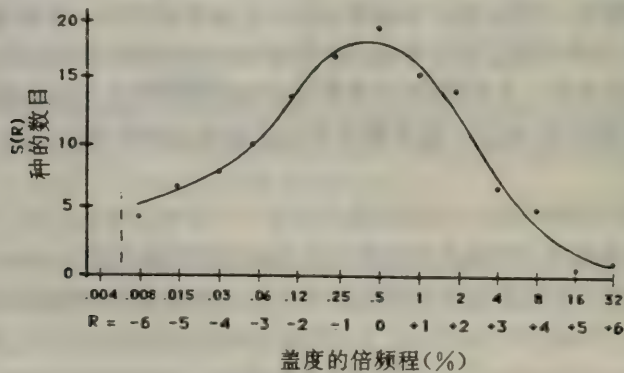


图 7.2 亚利桑那州,圣卡他利那山种类丰富的荒漠群落对数正态分布。配合的曲线由 $S(R) = 17.2e^{-(0.235R)^2}$ 给出, $S(R)$ 是距离众数倍频程 $R=0$ 为 R 的倍频程上种的数目

由于每个倍频程代表一个加倍的多度,众数右侧相继倍频程的 N_i/N_0 比将分别为 2, 4, 8, 16, 32 等等, R 将分别为 1, 2, 3, 4, 5 等等;众数左侧倍频程的 N_i/N_0 将依次为 0.5, 0.25,

0.125, 0.063, 0.031 等等, R 将分别为 $-1, -2, -3, -4, -5$ 等(见图 7.2)。

下面显示的是一种总结观测频度分布的方便方法, 在这个例子中, 数据组有 8 个倍频程(倍频程 4 是众数), 在第 R 个倍频程上有 5 个(R)种:

				观测的	
倍频程	N_i/N_0	R	R^2	$S(R)$	$\ln S(R)$
1	0.125	-3	9	$S(R=-3)$	$\ln S(R=-3)$
2	0.25	-2	4	$S(R=-2)$	$\ln S(R=-2)$
3	0.5	-1	1	$S(R=-1)$	$\ln S(R=-1)$
4	1	0	0	$S(R=0)$	$\ln S(R=0)$
5	2	1	1	$S(R=1)$	$\ln S(R=1)$
6	4	2	4	$S(R=2)$	$\ln S(R=2)$
7	8	3	9	$S(R=3)$	$\ln S(R=3)$
8	16	4	16	$S(R=4)$	$\ln S(R=4)$

R^2 这一列的给出是因为它出现在方程(7.1)中, $S(R)$ 的对数是用在步骤 2 中对对数正态分布的参数估测。

图 7.2 的对数正态分布曲线沿 x 或“倍频程”轴的两个方向趋向于无穷, 也就是说, 在理论上观测到的所有种类的数目都是可以利用的(Preston 1948)。对于小样本, 仅仅曲线的右侧是可见的, 随着样本规模的增加, 也就是说在这个样本中一些稀有种可能被包括进来了, 分布的左侧也就变得可见了(即曲线的众数将向右移)。图 7.2 中的竖直线, 它将这条曲线的取样部分从未取样部分分开, 称之为“垂线”, 这种取样方法的使用要求测定的值不能低于 0.004~0.008%(Whittaker 1965)。

在构造频度图时, 如果一观测多度恰巧落在两个倍频程的分界线上, 按常规, 这个多度的 1/2 将被划分在那个较低倍频程上, 另一半分到那个较高的倍频程上。例如, 使用每个种的个体数目作为多度的测定, 如果 10 个种的每个种都有 8 个个体, 那么 5 个种将被放在 4 到 8 个个体的倍频程内, 剩余的 5 个种将被放在 8 到 16 个个体的倍频程中(这个问题将在 7.3 节中的例子阐述)。

步骤 2. 参数的估计。正态对数分布由两个参数 S_0 和 a 表示它的特征(公式 7.1), 我们将阐述估计这些参数值的几个简单方法。但我们也应当让学生们知道非线性回归也可用来估计这些参数(见 Sokal 和 Rohlf 1981)。

参数 a 的一个近似由下式给出。

$$a = \sqrt{\frac{\ln[S(0)/S(R_{max})]}{R_{max}^2}} \tag{7.3}$$

$S(0)$ 是在众数倍频程上种的观察数目, $S(R_{max})$ 是距离众数最远的一倍频程(由 R_{max} 表示)种的观察数目。在这个例子中的步骤 1, $R_{max} = 4$ 。当观测的数据由众数数起至少包括 3 或 4 个倍频

程时,公式(7.3)才能表现最好。同时,在众数的两个方向上,如果最远的倍频程相同的话,比如-5和+5,那么a的值应用方程(7.3)对 $S(R_{\max})$ 的两个值计算并取平均值作为a值。

在生态学的大多数样本中,参数a大约为0.2,因而,它曾引出了关于它为什么是这样的许多假设,但May(1975)指出,这条“规律”事实上是对数正态分布数学成分的一个产物。由于群落中种的总数从20到10000变化,参数a也将在0.3到0.13间变化(这被认为是潜在的分

布遵循所谓的“典型对数正态分布”,见7.6节)。

参数 S_0 的估计可由在众数倍频程上的种数目的观测值或由下式获得。

$$S_0 = e^{(\overline{\ln S(R)} + a^2 \bar{R}^2)} \quad (7.4)$$

$\overline{\ln S(R)}$ 是每个倍频程上种观测数目对数的平均值,a由方程(7.3)估测。 \bar{R}^2 是 R_i^2 的平均值。

传统上,对数正态分布是由肉眼来配合的(Vandermeer 1981),但方程(7.3)和(7.4)提供了参数的合理统计估计量。程序LOGNORM.BAS允许用不同的a和 S_0 反复迭代到公式7.1中,直到在这些倍频程上种的观测值(步骤1)和期望值(步骤3)之间的离差最小。

步骤3.期望频度。使用 S_0 和a的估测值,应用公式(7.1)计算对数正态分布的期望值。同时用卡方检验此模型与观测频度的吻合性,自由度为倍频程数减2,由于我们仅仅是想获得一个近似的配合,这个卡方统计量应作为参数选择的一个指导,而不是作为一个正式的统计检验。

7.3 例子:路中的昆虫

下面的昆虫计数数据来自草原群落的路中,是通过组合14个扫网中的捕获物而获得的

(X):	1	2	3	4	5	6	7	9	10	11	21	28	23	120
(f):	32	8	9	2	3	3	3	2	1	2	1	1	1	1

X是每个种个体的数目,f是在每个X级别上种的频度。个体的总数等于389,种的总数为69。

配合对数正态分布

步骤1.观测频度。表7.1给出了每个倍频程种的数目,并将这些值绘制在图7.3中。注意第2个倍频程,它有20个种,是众数倍频程,同时也注意多度级落在连续倍频程分界上的种在那些倍频程上是怎样被分离的:

倍频程1 (0-1): $32/2=16$

倍频程2 (1-2): $16(\text{从倍频程1而来})+8/2=20$

倍频程3 (2-4): $4(\text{从倍频程2来})+9+2/2=14$

倍频程4 (4-8): $1(\text{由倍频程3来})+3+3+3+0=10$

步骤2.参数的估计。a的一个估计值是来自公式(7.3),在 $R_{\max}=6, S_0=20, S(R_{\max})=1$ 和 $R_{\max}^2=36$ 时

$$a = \sqrt{\frac{\ln(20/1)}{36}} = 0.29$$

由公式7.4

$$\overline{\ln S(R)} = (2.77 + \dots + 0)/8 = 1.63$$

$$R_2 = (1 + \dots + 36)/8 = 11.5$$

所以

$$S_0 = e [1.63 + (0.29)^2(11.5)] = 13.2$$

使用 BASIC 程序 LOGNORM. BAS 的参数估计值是 $a=0.305, S_0=15.9$, 结果总结在表 7.2 中。

表 7.1 以倍频程表示的路中昆虫计数数据

倍频程	每个种个 体的数目	R	R ²	观测的 S(R)	lnS(R)
1	0-1	-1	1	16	2.77
2	1-2	0	0	20	3.00
3	2-4	+1	1	14	2.64
4	4-8	+2	4	10	2.30
5	8-16	+3	9	5	1.61
6	16-32	+4	16	2	0.69
7	32-64	+5	25	1	0.00
8	64-128	+6	36	1	0.00

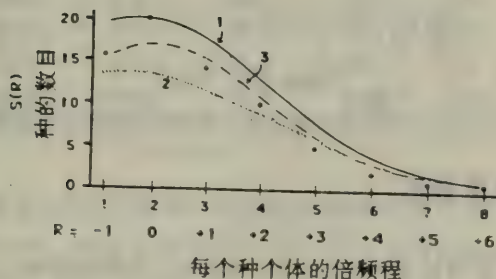


图 7.3 以倍频程表示昆虫计数数据频度分布图: 黑点表示观测的数据, 配合的对数正态模型如曲线 1 ($a=0.29, S_0=20$)、曲线 2 ($a=0.29, S_0=13.2$) 和曲线 3 ($a=0.305, S_0=15.9$) 所示。

步骤 3. 期望频度。 使用 $a=0.29, S_0=20$ 和 13.2 利用公式 (7.1) 计算期望频度。例如在 $S_0=20$ 时, 第一个倍频程上的 $S(R)$ 为:

$$S(-1) = 20e^{-(0.29)^2(-1)^2} = 18.4$$

注意: 三条对数正态分布曲线 (由三组参数给出) 频度期望总数分别为 89、59 和 69, 而观测的种的数目是 69, 为什么是这样呢? 回忆一下对数正态曲线下的面积是代表可以观测到的种的理论数目 S^* , 它可由下式计算。

$$S^* = 1.77(S_0/a) \quad (7.5)$$

对于前面给出的三组参数估计值, S^* 分别等于 122, 80 和 92。当检验图 7.3 中的三个对数正态分布图时这就更为明显, 曲线 1 下面的面积最大, 曲线 2 最小, 曲线 3 居中, 这样再回到表 7.2 的频度期望上, 我们可以看到第三组参数估计量(即反复迭代的)可能最适合这些数据, 而且大约有 23 种在此群落没被观测到。当然, 这只是为了叙述方便而采用的一个简单的数据组, 当配合对数正态时, 大样本是十分重要的。

表 7.2 观测的昆虫计数数据的对数正态模型的适合性检验, 并给出了三组参数估计值(见书)。

		a = 0.29			0.29			0.305		
		S ₀ = 20			13.2			15.9		
倍频程	R	观测 S(R)	期望 S(R)	χ ²	期望 S(R)	χ ²		期望 S(R)	χ ²	
1	-1	16	18.4	0.31	12.2	1.19		14.5	0.16	
2	0	20	20.0	0.00	13.2	3.45		15.9	1.06	
3	+1	14	18.4	1.05	12.2	0.27		14.5	0.02	
4	+2	10	14.3	1.31	9.5	0.03		11.0	0.08	
5	+3	5	9.5	2.10	6.3	0.25		6.9	0.52	
6	+4	2	5.3	2.04	3.5	0.64		3.6	0.70	
7	+5	1	2.5	0.90	1.7	0.26		1.6	0.20	
8	+6	1	1.0	0.00	0.7	0.17		1.1	0.02	
		69	89.4	7.71	59.3	6.26		69.1	2.76	

7.4 荒漠群落的盖度数据

这个例子中, 我们使用的是 Whittaker (1965) 在亚利桑那州, 圣卡他利那山低坡度种丰富的荒漠群落获得的数据。这些数据在图 7.2 中给出, Whittaker 使用种的百分盖度作为多度的测定。并以倍频程形式安排数据如下:

R:	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	+6
S(R):	4	7	8	10	14	17	19	15	14	7	5	0	1

$S(R)$ 是距离众数倍频程为 R 的每个倍频程上种的观测数目, 从图 7.2 可以看出, $R = -6$ 这个倍频程上盖度在 0.008—0.015 之间等等。

使用程序 LOGNORM.BAS 结合初始值 $a = 0.247$ 和 $S_0 = 17.9$ [由公式 (7.4) 得] 反复迭代运算, 最低的卡方值在 $a = 0.235$, $S_0 = 17.2$ 处获得, Whittaker 1965 对 a 和 S_0 的估计值分别为 0.245 和 17.5, 他可能通过非线性回归获得。这个对数正态曲线绘制在图 7.2 中, 注意观测的种总数是 122, 而由公式 (7.5) 理论上可以观测到的数目 S^* 为 129。

7.5 例子:纽约种禽。

这个例子来自 Preston(1948)对纽约州西部种禽的普查数据的总结,下面的数字是每个种育种对的数目(以倍频程表示):

R:	-6	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
S(R):	1	1.5	6.5	8	9	9	12	6	9	11	4	3

S(R)代表距离众数(R=0)为 R 的倍频程上种的数目。

使用 LOGNORM. BAS 我们计算的最吻合估计值是在 $\chi^2=7.05, a=0.21, S_0=10$ 时。这些数据和对数正态曲线在图 7.4 中给出。再一次请注意一下 a 值是常常接近于观测到的值 0.2 的(步骤 2, 7.2 节),而观测到的种的数目是 80,理论上可以观测到的数目 S* 大约为 84[公式(7.5)]。

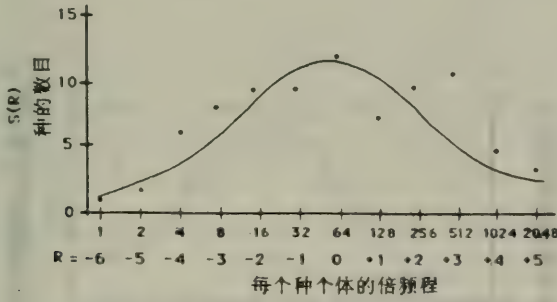


图 7.4 种禽普查数据(黑点)和配合的对数正态分布(a=0.21, S₀=10.0)

在 Preston(1948)的工作中,有许多数据可以利用以获得与程序 LOGNORM. BAS 计算的相类似结果。用程序 LOGNORM. BAS 获得的和 S₀ 估计值可以同 Preston 获得的那些相比较。

7.6 对数正态分布模型的附加讨论

在生态学中对数正态趋向于适合大量的例子,但还有些争执,即这是反映生物过程(如群落相互作用)呢? 还是仅仅与统计学中的大数定理有关(Minshall 等 1985)。

在 7.2 节里,我们讨论了这条感兴趣的“规律”,即对数正态分布中参数 a 通常是在 0.2 左右。另一个与对数正态分布有关的经验规律就是 Preston 的典型假设(Canonical Hypothesis)。Preston 用对数正态模型分析了许多组数据,以检验每个倍频程上种的数目与每个倍频程上这些种的个体总数之间的关系。图 7.5 中阐述了这种比较的一个例子。图 7.5 中的“种曲线”是每个倍频程上种的数目[S(R),公式(7.1)],而“个体”曲线是每个倍频程上个体的数目。I(R) 由下式获得

$$I(R) = S(R)N(R) \quad (7.6)$$

$N(R)$ 是第 R 个倍频程上每个种个体的数目,Prerton 以 R_{max} 即种曲线上种最丰盛种的倍频程与 R_N 即个体曲线的众数倍频程之间关系的形式比较了这两种曲线,它们的比为:

$$q = R_N/R_{max} \quad (7.7)$$

当 $q=1$ 时表明对数种曲线是典型的,如图 7.5 所描述的那样,这个经验规律趋向于可以应用到用这种方法分析的许多群落数据。

May(1975)提供了与对数正态分布有关的这两条经验规律的彻底分析。他指出在种和个体总数 S 和 N 与参数 a 和 q 之间存在一个基本关系,即当 S 的值从 20 到 10000 和 N 从 10S 到 107S 变化时,参数 a 和 q 就分别在 0.1~0.4 和 0.5~1.8 之间。May 指出这仅映了统计的一般原则和大样本之间的关系。

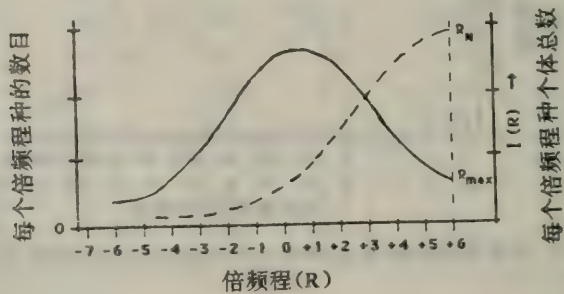


图 7.5 对数正态种曲线(实线)和与它相关的个体曲线

但 Sugihara 的研究表明,对于这个典型假设,生态群落并不是严格遵从象 Mayer 统计讨论给出的所有的解释那样,他画了许多群落的 S 对应于对数正态标准差 σ 图,结果表明,对于一给定的 q 值, S 和 σ 之间似乎存在一个独特的关系,他还发现自然群落非常接近对数正态典型形式($q=1$),他同时还提出了可能引起这种格局的一个生物学机制。

Minshall 等(1985)用对数正态模型作为一个经验方法测定了河流生态系统无脊椎动物群落的平衡状况。他们提议如果一个群落与对数正态的一致性越高,其稳定性程度也就越高,这样通过这个群落在不同季节与对数正态的一致性估计就可作为这个系统平衡性的一种测定。这当然是格局检验应用比较好的一个例子,不管对于这个模型提出的潜在假设真实与否。虽然这个方法在 Minshall 等的研究中表现的很好,但 Gray(1981)发现用类似的方法在试图检验河流的污染状况时失败了,因为出现的对于对数正态较好的配合是没有考虑到在群落组成方面的状况和变化。

7.7 总结和推荐

1. 在一个群落中当将每个种的相对多度按着种序由最丰盛到最不丰盛的顺序绘制成图时,特定的格局就出现了,如图 7.1 所示。

2. 生态学家提出了许多假设以试图解释观测到的多度关系,但发现相反的假设可以导致相同模型的产生,在某些情况下,不同的模型可以配合同一组数据,迄今,没发现一个关于多度关系的普遍模型。

3. 对于一个大量种的集合,相对多度趋向于遵从对数正态分布,也许仅仅是由于大数定理的结果,但大多数群落看上去是遵从对数正态分布的一种特定形式,即典型的形式,所以我们推荐在检验一个群落种多度的格局时应用对数正态模型。

4. 我们所提出的关于种多度关系这些模型和混乱难懂的东西是一些检验格局和趋势的有用工具,这些工具在产生关于群落组织的一些站得住脚的假设时是有用的。

第八章 多样性指数

在这一章中我们要描述几个多样性指数，这几个多样性指数可以用作群落中物种多度相互关系的特征。多样性由两个独立的成分构成：(1) 物种总数，和 (2) 均匀性（多度数据是如何在物种间分布的）。由于多样性指数往往试图把这两种成份掺在一起使之成为单一的数量值，因此涉及到对它们进行解释和正确使用时，常常导致争论和混乱。在本章中我们希望能阐明一些争论的焦点。

8.1 基本方法

在群落生态学中多年来生态学家们一直就物种的多样性问题进行激烈争论。实际上 Hurlbert (1971) 曾就此问题提出建议：可能多样性最好被描述为“无概念”，因为有许多语义上的、概念上的和技术上的问题与这个术语的使用有关联。涉及到多样性指数的使用，虽然列举出许多争论和值得注意的意见，但多样性指数仍在生态学家中广为流行。多样性指数同大多数方法一样是可靠的，它可以相对简单地获得一些基本知识，继而推进计算方法；对于获得决定性的观点，它有较强的挑战性。为此，我们对初学的学生的第一个要求就是弄清所有多样性指数的限制。

可以认为物种的多样性由两个成分组成。第一个成分是群落中物种的数量，生态学家通常称之为物种丰富度 (species richness)。第二个成分是物种的均匀性 (species evenness) 或平均性 (equitability)。均匀性是指物种多度（例如个体数量、生物量、盖度等）是怎样在各物种中分布的。例如：在一个由十个种组成的群落中，如果 90% 的个体属于一个种而剩余的 10% 在其他九个种间分布，则可以想像均匀性很低。另外，如果十个种都占有 10% 的总个体数，则可想而知均匀性达到最大位。

多年来为了使物种丰富度和均匀性数量化已提出了许多指数。这样的指数称之为丰富度指数 (richness indices) 和均匀性指数 (evenness indices)。试图把物种的丰富度和均匀性联系起来，使之成为单一值的指数就是我们所谈到的多样性指数 (diversity indices)。因此所有多样性指数的主要关键是这些指数试图将已发现的许多使群落结构特征化的变量：(1) 物种数量，(2) 种的相对多度 (均匀性)，和 (3) 均一性和取样面积大小联系在一起 (见 James 和 Rathbun 1981)。下面当我们为了计算多样性指数而描述各个步骤时将考虑这些问题。

8.2 步骤

8.2.1 丰富度指数

显然物种丰富度的一个即明确而又直接了当的系数是 S — 群落中物种的总数。然而，由于 S 取决于样本含量 (和寻找所花费的时间)，它被限定为可比较的指数 (Yapp 1979)。所以

提出了许多指数以测定独立于样本规模的物种丰富度。这些指数以 S 和观察到的个体总数 n (随样本规模的增加而增加) 之间的相互关系为基础。

下面列出了两个在历史上众所周知的丰富度指数：指数 1, Margalef (1958) 指数,

$$R1 = \frac{S - 1}{\ln(n)} \tag{8.1}$$

和指数 2, Menhinick (1964) 指数,

$$R2 = \frac{S}{\sqrt{n}} \tag{8.2}$$

这些指数有多大用处？为了回答这个问题，考虑应用 $R2$ (Menhinick 指数) 的实际结果。假定用 $R2$ 来测定在沿山坡的不同高度的三个取样地上鸟类群落的物种丰富度。在最低的取样地，设 $S=30$, $n=100$, 从而 $R2=3$ 。在斜坡中部的取样地 $S=15$, $n=25$ 则 $R2=3$ 。最后在最高取样地 $s=10$, $n=25$, 则 $R2=2$ 。基于这些 $R2$ 值我们能否得出结论说：除到达高海拔处以外，在 $R2$ 值无差异的山坡上从低海拔到中海拔处鸟类群落的物种丰富度最大？在得出结论以前重要的是弄清在应用 $R2$ 时要预先假定群落中 S 和 n 之间存在函数关系，即事实上给定为 $S=k \sqrt{n}$, 其中 k 为常数。这一点一定要考虑进去，否则 $R2$ 将随样本包括的 n 值不同而变化，结果群落不能进行比较。所以，在上面所举例子中，关于物种丰富度的结论取决于是否已实际给定 S 与 n 间存在的 $S=k \sqrt{n}$ 的函数关系，如已给定则 k 为常数。正如 Peet (1974) 指出的那样，如果这些假设未被考虑进去，则丰富度指数将随样本含量的变化以某种不可知的方式变化。最后我们要求只有把这些假设全部弄清楚才能把 $R1$ 和 $R2$ 用作丰富度指数。我们怀疑：在大多数情况下没有满足这些假设的优先判定式，也没有结论得出。

对于丰富度指数而言，一个可采用的方法是在等量的样本中做种数的直接计数。这不仅是一个非常简单的步骤，还可避免如上面所描述的那种类型的指数使用问题。在样本含量不相等的场合（可能是通常的场合）统计学中众所周知稀疏 (rarefaction) 法用来比较群落间的物种数 (Hurlbert 1971, Sanders 1968)。为了使用稀疏方法，我们假定群落间的样本含量偏差或取样差异可以由一些能应用于所有有关群落的基础取样模式所抵消。以下给出这种模式的例子。

Hurlbert (1971) 表明：可以期望在一个从总个体数为 N , 且个体分布于 S 种间的种群中

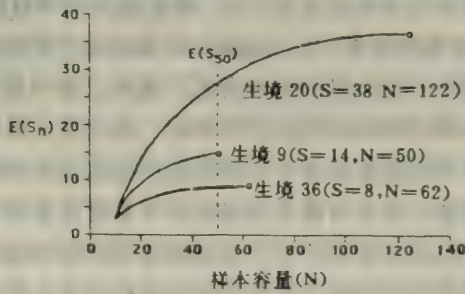


图 8.1 三个鸟类栖息地的稀疏曲线，示把物种数期望为样本含量的函数。

引自 James 和 Rathburn (1981)

抽取的个体数为 n 的样本 [记作 $E(Sn)$] 中种数为

$$E(S_n) = \sum_{i=1}^s \left\{ 1 - \left[\frac{\binom{N-n_i}{n}}{\binom{N}{n}} \right] \right\} \quad (8.3)$$

其中 n_i 是第 i 物种的个体数。简言之，方程式 (8.3) 把含量为 n 的随机样本中期望的种数计算为样本中包括每个物种的概率之和。

包括概率计算在内的关于怎样实际应用方程式 (8.3) 的问题已在计算机程序 RAREFR. BAS (见附光盘) 中给出；在这里要阐明的是作为丰富度测定的可能性。

可以在 James 和 Rathbun (1981) 的文章中找到应用的稀疏法的优秀范例，他们研究了在美国和加拿大广泛栖息的 37 种饲养鸟类的统计数字。我们以很高的评价向对这种方法的细节感兴趣的学生推荐这篇文章。图 8.1 示出他们文章的部分结果。在栖息地 20 (硬木—松树草甸) 通过观察得到的统计数字为总共 38 种 (S)，122 只鸟 (N)；这个点在图 8.1 中以一个圆弧代表。为了计算这个栖息地的稀疏曲线，他们使用方程 (8.3) 来计算当 120, 110, 100 等不同样本容量时物种的期望数，即 $E(S_n)$ 。这个过程在图 8.1 中以上部的曲线来表示 (注意 N 在这里是作为种群参数处理的)。对于 $S=14$, $N=50$ 的栖息地 9 (加拿大短针松—桦树林) 和 $S=8$, $N=62$ 的栖息地 36 (牧豆树—桤柳—拉瑞阿荒漠) 也采取了同样的处理步骤。现在可以用每一栖息地的稀疏曲线来抵消总鸟数的差异 (即分别为 $N=122$, 50 和 62)。James 和 Rathbun 把 $n=50$ 作为样本含量的标准，与做过统计的 37 个栖息地中最小样本含量相一致；这在图 8.1 中以 $n=50$ 处所绘垂直虚线来表示。在此样本含量我们可以就物种丰富度对这三个栖息地进行排序。具有期望种数为 26.9 的栖息地 20 丰富度最高，而期望种数为 7.8 的栖息地 36 丰富度最低。

我们同意 Hurlbert (1971) 与 Games 和 Rathbun (1981) 的观点，即当群落的样本含量不同时，在可用的简单的丰富度指数中，选择稀疏方法是最好的。然而，Peet (1974) 表明：对于拥有差异很大的物种数和相对多度的两个群落来说，稀疏方法可能预示两个群落都在小的样本含量处具有相同的物种数。为此当应用这种方法时，要假定所研究的群落在物种一个体相关关系方面不存在差异 (Peet 1974)。所以我们在推荐将稀疏法应用于物种计数的同时再次强调我们在前面说过的——弄清任何多样性方法的限度。

8.2.2 多样性指数

多样性指数把种的丰富度和均匀性综合在一起，使之成为一个数值。由于这个原因，Peet (1974) 把这些指数术语化为异质性指数 (heterogeneity indices) 可能在应用多样性指数中要克服的最大障碍就是说明这种单一统计量的实际含意。例如，在一些情况下一个给定的多样性指数值可以由种的丰富度和均匀性的不同组合中导出。换言之，具有低丰富度和高均匀性的群落与具有高丰富度和低均匀性的群落都可以得到相同的多样性指数值。接踵而来的问题是如果只给你一个多样性指数值，你不可能说出是种的丰富度还是均匀性更具相对重要性。

虽然存在着这样的问题，生态学家在研究中仍普遍使用多样性指数，且在应用中常常忽略了一些众所周知的问题。这里给出的概述旨在介绍一些较为流行的多样性指数，同时就其意义的说明 (如果可能) 方面进行强调。

确实有无限多个多样性指数 (Peet 1974) 这些指数的单位差异极大，难以比较而又混乱，且附带着许多要说明的问题。我们相信由 Hill (1973b) 提出的多样性数 (diversity numbers)

系列可能是生态学中最容易解释的。

Hill 的多样性数族的方程形式是

$$N_A = \sum_{i=1}^S (P_i)^{1/(1-A)} \quad (8.4)$$

其中 P_i 是属于第 i 个物种的个体 (或生物量等) 所占比值, 这个方程的推导由 Hill (1973b) 给出。Hill 列出的这些多样性数的第 0, 1, 2 阶 [即在方程 (8.4) 中 $A=0, 1, 2$] 正好符合三个最重要的多样性测定值 (注意: 见 Hill 1973b 当 $A=1$ 时方程 (8.4) 的解释)。Hill 的多样性数是:

$$\text{数 0: } N_0 = S \quad (8.5a)$$

其中 S 是种的总数,

$$\text{数 1: } N_1 = e^{H'} \quad (8.5b)$$

其中 H' 是香农指数 (Shannon's index) (定义如下), 和

$$\text{数 2: } N_2 = 1/\lambda \quad (8.5c)$$

其中 λ 是辛普森指数 (Simpson's index) (定义如下)。

这些以种数为单位的多样性数测度, Hill 称之为存在于样本中的物种有效数 (effective number of species)。这个物种有效数是比例多度在各物种间分布程度的测度。显然 N_0 是样本中所有种的数量 (不考虑其多度), N_2 是非常丰富的物种的数目, 而 N_1 测定样本中丰富的种的数目, (N_1 总是介于 N_0 和 N_2 之间。) 换言之, 物种的有效数是样本中种数的测度, 在此样本中每个种由其多度为权。

Hill 数间的差异仅在于样本在是包括还是忽略罕见种的趋势上 (Alatalo 和 Alatalo 1977)。作为举例考虑一个有 11 个种和 100 个个体的样本, 其中多度的分布是 90, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1。显然由于一个种非常丰富, 我们可以期望 N_2 接近于 1, 事实也正如此 ($N_2=1.23$) 当然 N_0 等于 11, N_1 是 1.74, 是一个介于 N_0 和 N_2 的值。

在直觉上 Hill 数作为多样性的指数对生态学家很有吸引力 (Peet 1974)。除对这些指数的解释中尚存在一些不清楚外 (Hill 1973b), 我们发现与其他可用的指数相比 Hill 数不那么混乱。进一步强调: Hill 数的单位是种, 且当种数增加时: (1) 罕见种的权变得更小 (回忆最低的数 N_0 , N_0 是样本中所有种的数目), (2) N_1 和 N_2 的取值变低, 因为这两个数在样本中是分别对丰富的种及非常丰富的种的测度。

在这一章中我们把多样性指数的范围限定于 N_1 和 N_2 。 N_1 和 N_2 作为多样性指数适合于讨论“异质性指数可以回答的任何问题” (Peet 1974)。还有, 许多流行的多样性指数只是 N_1 和 N_2 的变形。在下面进行的计算中我们专门使用计数资料, 即每个第 i 物种的个体数目 (n_i) 和取样个体的总数 (n)。然而, 如同 Hurlbert (1971) 和 Lyons (1981) 曾指出的那样: 大多数生态学家同意群落中物种的重要性应基于象生物量或生产力那样的量——如果可以得到这样的数据。这些量可以用来代替计数。但是, 重要的是要意识到当进行这样的代换时, 通常用来计算多样性测度值方差的方法就不再适用了。我们建议学生去参考 Lyons (1981) 的细节。

为了计算 Hill 的多样性数需要两个指数: 辛普森指数 λ 和香农指数 H' 。这两个指数描述如下。

多样性指数 1. 辛普森指数 λ . 辛普森 (1949) 提出的第一个应用于生态学中的多样性指数是

$$\lambda = \sum_{i=1}^S P_i^2 \quad (8.6)$$

其中 P_i 是第 i 个种的比例多度, 给定为

$$P_i = \frac{n_i}{N} \quad i = 1, 2, 3, \dots, S$$

其中 n_i 是第 i 个种的个体数, N 是总体中 S 个种的已知个体总数。变化于 0 到 1 之间的辛普森指数给出在总体中随机抽到属于同一物种的两个个体的概率。简言之: 如果属于同一物种的两个个体被抽出的概率高, 则群落样本的多样性就低。

方程 (8.6) 只应用于对所有群落成员进行过计数的有限群落, 即 $n=N$, 其中 n 是样本中个体总数, N 是总体的总个体数。由于我们在工作中通常所遇的是不可能对全部成员进行计数的无限总体, 所以辛普森 (1949) 导出从无限总体中取样的无偏估计量 ($\hat{\lambda}$) 为

$$\hat{\lambda} = \sum_{i=1}^S \frac{n_i(n_i - 1)}{n(n - 1)} \quad (8.7)$$

实际上, 当我们在方程 8.6 中以 n 代替 N 时, 方程 (8.6) 和 (8.7) 近似相等。

$\hat{\lambda}$ 的倒数是 Hill 的第二个多样性数 N_2 [见方程 (8.5c)]。许多研究者在理论上和实际工作中都多选择 N_2 , 选择 $\hat{\lambda}$ 则较少 (见 Alatalo 和 Alatalo 1977, Peet 1974, 和 Routledge 1979)。

多样性指数 2. 香农指数 H' . 在群落生态学中香农指数 (H') 可能是应用最广泛的指数。香农指数以信息论为基础 (Shannon 和 Weaver 1949), 就预测从含有 S 种和 N 个体的集合中随机选取的一个个体属于哪个种的问题而论, 香农指数是“不确定性”程度的一种平均测度。这个平均的不确定性随种的数量增加而增加, 也随种间个体的分布变得均匀而增加。所以 H' 有两个使之成为广为流行的物种多样性测度的性质: (1) 如果, 且只能当样本中只含有 1 个种时, $H' = 0$, (2) 只有当所有的 S 种均由相同的个体数来表示, 即多度相当均匀地分布时 H' 达到最大值。

使用自然对数 (\ln) 的香农函数的方程是

$$H' = - \sum_{i=1}^{S^*} (P_i \ln P_i) \quad (8.8)$$

其中 H' 是在一个由具已知比例多度为 $P_1, P_2, P_3, \dots, P_{S^*}$ 的 S^* 种组成的无限群落中每个种的平均不确定性。 S^* 和 P_i 是总体参数, 在实际应用中, 样本的 H' 估计为

$$\hat{H}' = - \sum_{i=1}^S \left[\left(\frac{n_i}{n} \right) \ln \left(\frac{n_i}{n} \right) \right] \quad (8.9)$$

其中 n_i 是 S 种样本中第 i 种的个体数, n 是样本中总个体数。方程 (8.9) 是生态学文献中使用最频繁的香农指数形式。但是, 这个估计量是有偏的, 因为在群落中种的总数 (S^*) 可能大于在样本中观察到的种数 (S)。幸运的是如果 n 增大则偏差将变小。

DeJong (1975) 证明: H' 与样本中种数的对数线性相关。这是由 Hill 的 N_1 给定的香农指数形式 [方程 (8.5b)], 其中单位是种。换言之, 如果种数都相同, 则 N_1 给定的种数会导出

同一样本的 H' 。这里所谈的可能通过简单的数量举例才能很好地理解, 给定三个分别由 100, 50 和 700 个个体所代表的种的群落样本, $H' = 1.05$ 而 $N1 = 2.87$ 。所以相等多度的 2.87 种给定 $H' = 1.05$ 。

8.2.3 均匀性指数

当样本中所有的种都同样多时, 可直观地看出均匀性指数应是最大值, 当种的相对多度与均匀性背道而驰时, 均匀性指数向 0 递减。Hurlbert (1971) 注意到均匀性指数具有这种性质——如果这些指数可由下面任何一个数学式来表示的话:

$$V' = \frac{D}{D_{\max}} \quad (8.10a)$$

或

$$V = \frac{D - D_{\min}}{D_{\max} - D_{\min}} \quad (8.10b)$$

其中 D 是某个观察到的多样性指数; D_{\min} 和 D_{\max} 分别是可以获得的 D 的最小值和最大值。我们在下面的讨论中还将提到这两个问题。

在用数量表示多样性的均匀性成分的尝试中提出了许多指数。我们仅限于描述其中的五种, 而其每一种都可由 Hill 数的比来表示。为了对均匀性指数做深入的讨论, 我们建议学生去研究 Alatalo (1981) 的文章。

均匀性指数 1 (E1), 可能生态学家应用最普遍的均匀指数是

$$E1 = \frac{H'}{\ln(S)} = \frac{\ln(N1)}{\ln(N0)} \quad (8.11)$$

这是熟悉的 Pielou (1975, 1977) 的 J' , 当样本中的每个种 (即 $\ln S$) 均具有一个个体那样相当均匀等, J' 表示与可得到的 H' 的最大值有关的 H' 。注意 E1 的形式是方程 (8.10a)。

均匀性指数 2 (E2). Sheldon (1969) 提出以 E1 的取幂形式作为均匀性指数:

$$E2 = \frac{e^{H'}}{S} = \frac{N1}{N0} \quad (8.12)$$

均匀性指数 3 (E3). 如果把 E2 写成方程 (8.10b) 的形式, 即减去最小值后, 即变成 Heip (1974) 提出的均匀性指数:

$$E3 = \frac{e^{H'} - 1}{S - 1} = \frac{N1 - 1}{N0 - 1} \quad (8.13)$$

均匀性指数 4 (E4). Hill (1973b) 提出把 $N2$ 对 $N1$ 的比值作为均匀性的指数:

$$E4 = \frac{1/\lambda}{e^{H'}} = \frac{N2}{N1} \quad (8.14)$$

这是非常丰富种与丰富种的数量比。回忆上述讨论: 随着群落多样性的减少, 即随着一个种趋向于占优势, 则 $N1$ 和 $N2$ 都趋近于 1。在这样的条件下 E4 的值向 1 收敛 (Peet 1974)。

均匀性指数 5 (E5). 如果把 E4 写成方程 (8.10b) 的形式, 则变为

$$E5 = \frac{(1/\lambda) - 1}{e^{H'} - 1} = \frac{N2 - 1}{N1 - 1} \quad (8.15)$$

E5 是周知的修改的 Hill 比值。Alatalo (1981) 表明: 当一个种在群落中变得越来越占优势地

位时 E5 接近于 0 (不象 E4 那样接近于 1)。这对均匀性指数显然是一个理想的性质,也是为什么 E5 比 E4 更容易被接受 (Alatalo 1987) 的原因。

均匀性指数应独立于样本中的种数。在直觉上似乎很有理由地认为在不考虑现存的种数情况下,均匀性指数应该是不变的。Peet (1974) 曾表明: J' [E1, 方程 (8.11)] 受种的丰富度强烈影响; 把一个罕见种 (低 S) 加入只含几个种的样本中, 则 E1 值的变化会很大。表 8.1 阐述了这种敏感性, 在表中把仅一个个体所代表的种加入到含有三个很有代表性的种的样本中。E2 和 E3 [方程 (8.12) 和 (8.13)] 象 E1 一样对种的丰富度也非常敏感。相反, E4 和 E5 相对不受种丰富度的影响。

表 8.1 由两个样本计算的均匀性指数, 样本 2 与样本 1 的差异仅在于加入了一个个体的新种 (Peet 1974 和 Alatalo 1981 修改)

样本	S	个体多度	E1	E2	E3	E4	E5
1	3	500, 300, 300	0.94	0.93	0.90	0.94	0.91
2	4	500, 299, 200, 1	0.75	0.71	0.61	0.94	0.90

当应用 E1, E2 和 E3 时, 应知道群落中现存种的实际数量 (S^*)。然而在实际应用中 S^* 通常由样本中现存种的数目 S 来估计。这通常导致对 S^* 的低估; 因此在估计指数时要引入数量偏差 (Pielou 1977)。Peet (1975) 注意到当每个种的偏差不妨碍对一般模式做出生态学解释的时候, 样本变差的高敏感性可能使指数毫无用处。另一方面, E4 和 E5 保持了样本变差的相对稳定 (例如罕见种的出现, 表 8.1), 从而趋向独立于样本含量。这是因为 E4 和 E5 是以比值计算的, 在这里 S 即是分子又是分母, 故而可有效地抵消样本中种数的作用。

群落的相对均匀度可通过观察稀疏曲线的峭度来进行比较 (8.2.1 节)。较高的均匀度等于较陡的稀疏曲线。在图 8.1 中所示的例子内栖息地 20 的均匀度最高而栖息地 36 最低。

8.3 例子：荒漠中蜥蜴的多样性。

我们假定：一个生态学家在 1 公顷的荒漠上数蜥蜴，他找到了属于 6 个种的共 36 只个体 (表 8.2)。利用这些数据我们将对不同种的丰富度、多样性和均匀度指数的计算进行阐述。

表 8.2 在 1 公顷样地上对 6 个蜥蜴种分别计出的个体数

蜥蜴种	个体数
Cnemidophorus tessellatus	3
cnemidophorus tigris	15
Crotophytus wislizenii	1
Holbrookia maculata	1
Phrynosoma cornutum	10
Sceloporus magister	2
	32

8.3.1 丰富度指数

丰富度指数 1. Margalef 指数 [方程 (8.1)]:

$$R1 = (6 - 1)/\ln(32) = 5/3.47 = 1.44$$

丰富度指数 2 (R2). Menhinick 指数 [方程 (8.2)]:

$$R2 = 6/\sqrt{32} = 6/5.66 = 1.06$$

8.3.2 多样性指数

多样性指数 1. 辛普森指数 [方程 (8.7)], 随机选取的两个个体属于同一物种的概率为:

$$\begin{aligned} \lambda &= [(3)(2) + (15)(14) + (1)(0) + (1)(0) + (10)(9) + (2)(1)] / [(32)(31)] \\ &= (6 + 210 + 0 + 0 + 90 + 2) / 992 = 308 / 992 = 0.31 \end{aligned}$$

Hill 的第二个多样性数 [方程 (8.5c)], 即非常丰富种的数目是

$$N2 = 1/0.31 = 3.22$$

多样性指数 2. 香农指数 [方程 (8.10)], 在预测从样本中随机选取的个体属于哪一个种的过程中不确定性的平均程度是

$$\begin{aligned} H' &= -[(3/32)\ln(3/32) + (15/32)\ln(15/32) \\ &\quad + (1/32)\ln(1/32) + (1/32)\ln(1/32) + (10/32)\ln(10/32) + (2/32)\ln(2/32)] \\ &= -[(-0.223) + (-0.355) + (-0.107) + (-0.107) + (-0.364) + (-0.172)] \\ &= -(-1.33) = 1.33 \end{aligned}$$

Hill 的第一个多样性数 [方程 (8.5b)], 丰富种的数目为

$$N1 = e^{1.33} = 3.78$$

和所期望的一样, $N1$ 介于 $N0$ (6, 种的总数) 和 $N2$ (3.22) 之间。

8.3.3 均匀性指数

利用 $S=6$, 均匀性指数可按下面的方式计算 [见方程 (8.11-8.15)]:

$$E1 = \ln(3.78)/\ln(6) = 1.33/1.79 = 0.74$$

$$E2 = 3.78/6 = 0.63$$

$$E3 = (3.78 - 1)/(6 - 1) = 2.78/5 = 0.56$$

$$E4 = 3.22/3.78 = 0.58$$

$$E5 = (3.22 - 1)/(3.78 - 1) = 2.22/2.78 = 0.80$$

8.4 例子: 佛罗里达河口湾的鱼

Livingston (1976) 在佛罗里达北部河口湾对鱼的季节性波动进行了研究, 该项研究所报导的捕捞数据在此用来比较种的丰富度、多样性和均匀性指数 (表 8.4)。BASIC 程序 SPDI VERS. BAS (在书内所附磁盘上) 被用来进行计算。

表 8.3 Livingston (1976) 取得的佛罗里达北部的河口湾中鱼类捕捞数据。
这是用拖网于 1972 年 12 月和 1973 年 6 月在 24 小时内捕捞的结果

鱼种	12 月	6 月
<i>Dasyatis sabina</i>	10	3
<i>Ophichthus gomesi</i>	3	0
<i>Dorosoma petenense</i>	1	1
<i>Anchoa mitchilli</i>	278	377
<i>Arius felis</i>	1	21
<i>Hippocampus erectus</i>	1	0
<i>Eucinostomus gula</i>	3	0
<i>Lagodon rhomboides</i>	1	0
<i>Bairdiella chrysura</i>	21	0
<i>Cynoscion arenarius</i>	19	30
<i>Cynoscion nebulosus</i>	7	0
<i>Leiostomus xanthurus</i>	18	238
<i>Menticirrhus americanus</i>	217	4
<i>Micropogon undulatus</i>	40	279
<i>Orthopristis chrysoptera</i>	0	1
<i>Peprilus paru</i>	1	0
<i>Prionotus tribulus</i>	35	0
<i>Prionotus tribulus</i>	25	0
<i>Etropus crossotus</i>	29	0
<i>Symphurus plagiusa</i>	50	3
<i>Monacanthus hispidus</i>	1	0
<i>Bagre marinus</i>	0	1
<i>Porichthys porosissimus</i>	0	4
<i>Polydactylus octonemus</i>	0	590
<i>Paralichthys lethostigma</i>	0	5
<i>Trinectes maculatus</i>	0	5
个体总数 N=761		1562
种总数 S=20		15

从十二月到六月的捕捞物中丰富度指数 R_1 和 R_2 下降。这反映种数的减少($N_0=20$ 到 $N_0=15$)和在六月捕捞物中个体总数的大幅度增加,这些丰富度值在此计算出的目的在于举例;所

表 8.4 在佛罗里达北部河口湾中两个取样时间鱼种多样性指数

指数	1972 年 12 月	1973 年 6 月
丰富度		
N_0	20.00	15.00
R_1	2.86	1.90
R_2	0.72	0.38
多样性		
λ	0.23	0.26
H'	1.91	1.54
N_1	6.78	4.96
N_2	4.39	3.90
均匀性		
E_1	0.64	0.57
E_2	0.34	0.31
E_3	0.30	0.26
E_4	0.65	0.83
E_5	0.59	0.79

以我们将不再做进一步的解释。相反,我们要涉及到图 8.2 中所示数据的稀疏曲线。以 761 为样本含量,十二月捕捞量的期望种数 $E(S_{761})$ 为 20,而六月捕捞量的 $E(S_{761})$ 是 13。所以

我们根据样本含量存在的差异所做比较推断：六月捕捞时种的丰富度下降。

从十二月到六月所有的多样性指数都下降（当然 λ 除外，因为 λ 的值随多样性减少而增加）。N1和N2是以种为单位的且最容易解释。这些数字指明在六月的捕捞中罕见种的增加占优势。六月的捕捞相当清楚：其中测度非常丰富的种的N2是3.9，而事实上，四个种占丰富度的95%（表8.3）。测度丰富种的N1在六月捕捞中是4.69，这个值将总是大于N2的值。

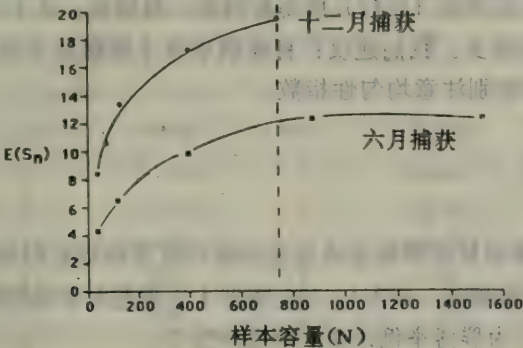


图 8.2 于十二月和六月在佛罗里达北部河口湾的捕捞物中期望物种数的稀疏曲线，数据取自 Livingston (1976)

均匀性指数较难以解释。我们把E1, E2和E3看作限定值，因为这些指数对样本中的种数极为敏感。从十二月到六月，样本的Hill比率（E4）和修改的Hill比率（E5）均增加。在此情况下，这种增加似乎与出现的15个种中的4个共优种和与其余的11种相比较这4个种非常相似的个体数有关（表8.3）。

8.5 例子：森林树木的多样性

表 8.5 1956 和 1975 年于哈佛林中以平方米为单位的胸高总断面积。摘自 Zahh (1977)

种	1956	1975
Quercus rubra	15.158	27.554
Betula papyrifera	8.601	11.845
Fraxinus americana	3.764	1.466
Acer rubrum	2.646	4.974
Pinus strobus	2.515	2.725
Betula lenta	2.240	2.210
Pinus resinosa	1.292	1.125
Hicoria ovata	0.622	0.680
Tsuga canadensis	0.527	0.808
Quercus alba	0.463	0.052
Acer saccharum	0.302	0.549
Quercus velutina	0.198	0.098
Pinus sylvestris	0.055	0.000
Betula lutea	0.024	0.030
Ostrya virginiana	0.012	0.009
Tilia americana	0.012	0.000
Populus grandidentata	0.006	0.000
Prunus serotina	0.000	0.012
Acer pennsylvanicum	0.000	0.021
Castanea dentata	0.000	0.009
总断面积 N=	38.438	54.166
种总数 S=	17	17

Zahl (1977) 在马萨诸赛州皮特色姆 (Petersham) 的哈佛研究林中于 1956, 1960, 1966, 1969 和 1975 年对 20 株树的总树干断面积数据进行了搜集。他的目的是研究森林再生过程中的多样性变化格局, 因为那片森林于 1918 年被皆伐, 此后未受扰动。在此例中我们只选取他于 1956 和 1975 两年中获得的数据。

应用 BASIC 程序 SPDIVERS. BAS, 计算出并在表 8.6 中概括出种的丰富度、多样性和均匀性指数。这些结果导出与 Zahl (1977) 所得相同的一般结论: 从 1956 到 1975 种的丰富度、多样性和均匀性都有所减少。我们建议: 将此例中各个指数的表现与前所举各例 (8.4 节) 中的指数进行比较, 要特别注意均匀性指数。

8.6 多样性的附加讨论

Peet (1974) 根据多样性指数对群落组成变化的相对敏感性把它们划分成两类: (1) 类型 I — 那些对稀有种的变化最敏感的, 如 N1; (2) 类型 II — 那些对普通种的变化最敏感的, 如 N2。在表 8.7 中我们利用人为群落举例说明了这些敏感性。

考虑 A、B 和 C 三个群落, 各有 100000 个个体分布于三个普通种和数目不同的稀有种中。群落 A 由三个普通种和 200 个稀有种组成, 每一稀有种占总个体数的 1% (即 200 个)。当这些稀有种变成只相当于总个体数的 0.01% (即 20, 群落 B) 时, N2 的值相对不受影响, 而 N1 的值增加到 151% (表 8.7)。N1 的这种巨大变化是类型 I 指数的典型 (Peat 1974)。在群落 C 中, 3 个普通种的相对量变了 (稀有种仍保持各为 20 个个体), 并重新计算其指数。如预期结果相同, N1 和 N2 都发生变化的同时, N2 (类型 II 系数) 受影响最大, 减少了 33%, 而 N1 只减少了 15%。

表 8.6 在哈佛林中两个年份的, 以 20 株树的树干总断面积为根据的种多样性指数。这些结果根据表 8.5 中给出的数据求出。

指数	1956	1975
丰富度		
N0	17.00	17.00
R1	4.38	4.01
R2	2.74	2.31
多样性		
λ	0.21	0.31
H'	1.84	1.55
N1	6.27	4.71
N2	4.79	3.25
均匀性		
E1	0.65	0.55
E2	0.37	0.28
E3	0.33	0.23
E4	0.76	0.69
E5	0.72	0.61

还需讨论的一个题目是与估计一个无限大总体的多样性有关系的一些问题。完整普查通常是不可能的, 但又必须取样, 这就难免有误差。通常所研究的有机体占据着空间的连续区, 故

而必须使用一些人为的取样单位(SU)(如:植物群落的样方或鱼类群落的拖网)。因此,结果必然受象种群的空间分布和SV的大小与形状这样一些事件影响。Zahl(1977)指出:任何多样性指数的取样分布都将是所有这些因子的函数。

表 8.7 多样性指数比较,示对稀有种和常见种变化的敏感性(Spp)。每一人为群落有 N=100000 个体,三个常见种和不同数目的稀有种

群落							
A				B			
指数	Spp.	个体数		Spp.	个体数	Spp.	个体数
1	1	20,000		1	20,000	1	10,000
2	2	20,000		2	20,000	2	10,000
3	3	20,000		3	20,000	3	40,000
4	4	200		4	20	4	20
5	5	200		5	20	5	20
⋮	⋮	⋮		⋮	⋮	⋮	⋮
	203	200		2,003	20	2,003	20
S	203			2,003			2,003
λ	0.12	—	0%	→	0.12	—	+50% → 0.18
H'	3.45	—	+27%	→	4.37	—	-3% → 4.23
N1	31.55	—	+151%	→	79.24	—	-15% → 68.99
N2	8.12	—	+2%	→	8.28	—	-33% → 5.55

Pielou (1966, 1975) 指出一个旨在估计多样性及其取样误差的有趣过程,以求解决这些取样问题。Pielou 的方法主要是众所周知的合并样方法 (Pooled-quadrat method), 包括在群落中设置 q 个随机样方 (SUs), 按随机顺序整理这 q 个样方, 先计算第一个样方的多样性, 然后算第一个样方加第二个样方的, 余类推, 逐次合并样方的观测值。累积样方计算出的指数值最初会增加, 但随着继续合并则趋于在一定水平上稳定下来。稳定后的系数值即可取作群落多样性的估计值。

最近, 几个作者发展出可以替代 Pielou 方法的方法。Zahl (1977) 介绍了用于估计本章中出现的任何多样性指数的“大折刀法”(jackknife method)。Zahl 的方法可以获得这些指数标准偏差的估计值, 这些估计值必须进行假设检验和估计置信区间 (Heltshe 和 Dicanzio 1985)。Heltshe 和 Forrester (1983a, 1985) 表明: 在一定条件, Zahl 的大折刀估计值和 Pielou 的合并样估计值是相等的。与此问题还有关系的是 Heltshe 和 Forrester (1983b) 应用大折刀法估计 S* 一群落中种的总数。由于大折刀法的细节超出了此书的范围, 我们建议同学们去查阅前面的参考文献。

群落中总种数 S* 的估计 [来自对数正态模型方程 (7.5)], 也被用作丰富度指数, 但不象方程 (8.1) 和 (8.2) 那样, 它要求种丰富度数据在统计学上符合对数正态频数分布。虽然计算 S* 需要做许多更进一步努力 (在数据要求和计算中), 我们仍把 S* 作为优于 R1 和 R2 的

丰富度指数来介绍。

8.7 总结和推荐

1. 弄清多样性测度的限度。这包括种的丰富度指数、稀疏模型、多样性指数和均匀度指数。这些测定值易于计算，但通常难以解释（8.1节）。

2. 除非有有力的，符合函数相关关系和在事实上接受假设的判断在握以外，不要使用种的丰富度指数 $R1$ 和 $R2$ （8.2节）。我们认为：在实际应用中很少出现这种情况。

3. 如果样本含量相等，我们建议用种的直接计数来比较群落间的丰富度。在样本含量不同的地方，我们建议使用稀疏模型。然而重要的是作为基础的种与个体间相关关系要相同（8.2.1节）。

4. 来自对数正态模型的 S^* 在应用上作为种的丰富度指数来介绍。

5. 我们把 Hill 的多样性数 $N1$ 和 $N2$ 在应用上作为物种多样性的测度来介绍。在几项研究中都表明 $N1$ 和 $N2$ 较其他多样性指数更容易解释，同时因以种数为单位而引人注目。

6. 作为均匀性的测度，我们介绍使用修改的 Hill 比率 $E5$ ，因为 $E5$ 最不模棱两可和最易解释。 $E5$ 不需要估计群落中的种数，它受样本含量的影响（8.2.3节）。

第四部分 种的亲合性

第九章 背景

生态学的群落是由许多共存的种构成的。一些群落可能有很多种(如热带森林);另一些则可能只有少数几种(如一条受污染的河流)。在第七和第八章中我们描述了一些以经验为根据的,确定群落中物种总数与一些种的丰富度测度(如总数)之间相关关系的模型。在本书的这一部分中,我们把注意力集中在检验共存种间的亲合性。共存种如何利用共同的资源?

例如,考虑一个物种丰富的湖中有四个大小相近的优势鱼种。它们直接竞争食物和空间吗?是否有些种专门在表水中取食,而另一些种在底部取食?当我们在空间上确定了种A的位置时,在那里,通常我们是否不可能找不到种B?就广义的观点而论,我们可以把这样种间相互作用定义为物种之间亲合性(affinity)的程度。

亲合性的测度之一是物种利用共同资源的重叠程度。用术语将此重叠定义为被其他物种分享的物种生态龛的各个部分。生态龛的研究是以象食性、微生物选择性,和活动的时间性(如取食)之类的属性为根据的。重叠的测度将在第10章中论及。

在第11章中我们将讨论种间关联(interspecific association)这一论题。在此例中,我们仅涉及测度在同一地点一块找到两个种的频度有多大的问题,检验这一共存种间亲合性(或缺乏亲合性)的方法是检查种的出现[在一系列取样单位(SUs)中]是否大于或小于它们处于独立情况下的期望值。如果确定了种间关联是正的或负的,我们就可以用指数来测度这种关联的强度。

种间关联取決的数据全在于物种出现或未出现。如果样本含有物种多度的数量测定值,我们就可以确定物种间在多度上的协变。这可导致对物种亲合性的问题作出回答。例如,当其他种的多度增加时,如果一个种的多度总是减少,则存在某类原因导致负的相互作用吧?种间协变的测度在第12章中讨论。

这些方法的每一种都旨在帮助生态学家确定物种相互作用的形式。当然,通过这种确定无法推断形成这类作用形式的潜在原因,尽管我们希望这种形式的确定将直接导出可检验的假设。

9.1 矩阵展示

Cattell(1952)指出:从两个独特的着眼点可以研究生态学数据矩阵:(1)下降列(SUs)或(2)交叉行(种)。根据这些选出的部分,一定的近似性(resemblance)测度是可用的。这些近似性

函数的分类由图 9.1 给出。重要的是要意识到选择合适的函数与这样的事实有关,即在生态数据矩阵中,我们认为种(行)是彼此依赖的,而 SU_s (列)是独立的样本(Legendre 和 Legendre 1983)。

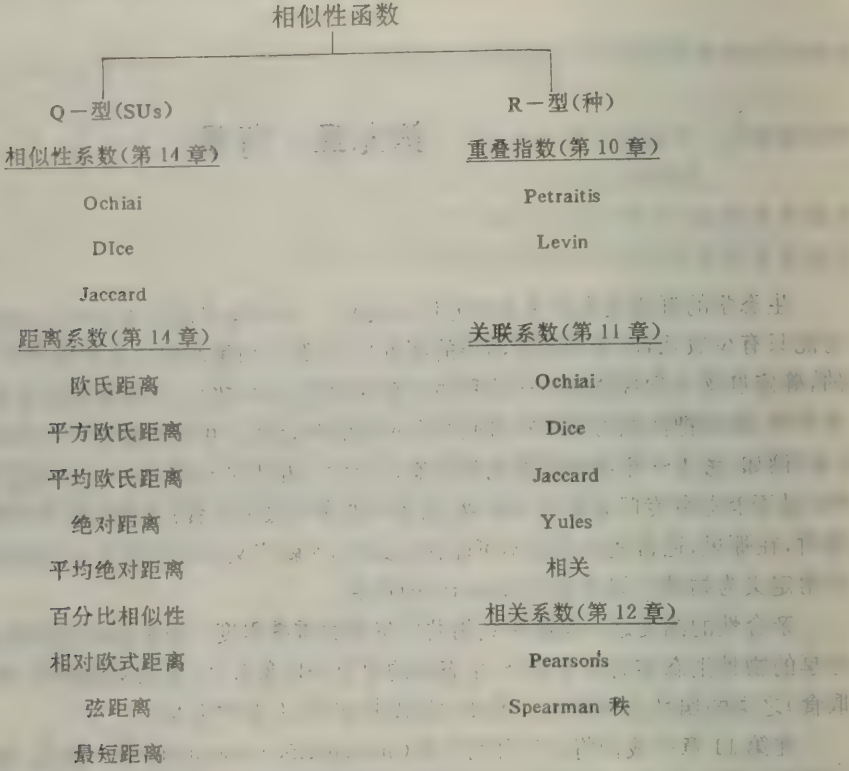


图 9.1 可应用 Q-型分析(相似性,距离)和 R-型分析(重叠,关联, 相关)的近似性函数

		取样单位										
		1	2	3	4	5	N
种	a											
	b											
	c											
	.											
	e											
环境因子	w											
	x											
	.											
	z											

图 9.2 阴影区指明测度物种亲合性的生态数据矩阵的形态。令人感兴趣的是与取样单位交叉的种的出现与否或多度。

通过对背景的讨论,同学们可以意识到在第 10-12 章中我们的兴趣在于种的亲合性,即根据与生态数据矩阵行交叉的数据来测度成对种的近似性(图 9.2)。生态学家将此称为 R-型分析(Legendre 和 Legendre 1983)。

R-型近似性函数(图 9.1)被分成重叠指数(第 10 章)、关联系数(第 11 章)和协变系数

(第 12 章)。这些 R-型指数测度物种之间亲合性依存关系或强度。直观地看来这是有道理的，因为我们所测度的是一块儿出现于 SUs 中的物种的近似性。

Q-型近似性函数根据取样单位中物种组成(既下降列)测度 SUs 间的相似性或相异性。再者，这种术语是有意义的，因为我们所比较的是独立于样本的类似或相异的程度。本书的第五到第七部分涉及大量的 Q-型分析问题。Q-型近似性函数在第 14 章中讨论。图 13.2 给出了 R-型和 Q-型方法的图解。

而图 9.1 所示的系统看起来是直观的，但在许多的情况下，在生态学文献中 R-型和 Q-型指数被互换了。例如，我们在第 11 章中描述的关联系数还被用来测度 SUs 之间的相似(第 14 章)。由于这些特殊系数的某些特性，这样应用是可接受的； 但用象 相关系数这样的 R-型系数去测度 R-型的近似性则不能同意(Qrloci 1972,1978)。

9.2 选择的文献

已在广泛的生物群落中对生态龛重叠、种间关联和协变进行了测定(表 9.1)。

表 9.1 为了举出研究物种亲合性例子而选出的文献,这些研究包括龛重叠(SN0),种间关联(SA),和种间协变(SC)

地 点	群 落	方 法	参 考 书
南达科他	草原鸟类	SNO	Garcat 和 Steinhorst 1976
佛罗里达	浅水底栖生物	SNO	Livingston 等人 1976
St. Croix	礁石鱼类	SNO	Gladfeltor 和 Gohpson 1983
加勒比	蝴蝶鱼	SNO	Findley 和 Findley 1985
加利福尼亚	荒漠灌丛	SA	Went 1942
伊利湖	鱼类	SA	Nash 1950
特里尼达	森林	SA	Greig-Smith 1952b
科罗拉多	普列利草原	SA	Cook 和 Hurst 1963
威斯康星	森林草本植物	SA	Smith 和 Cottam 1967
佛罗里达	潜叶虫	SA	Bultman 和 Facth 1985
英格兰	放牧场	SC	Kershaw 1967
英格兰	白垩土草原	SC	Greig-Smith 1983
尤他	底栖动物	SC	Drake 1984

第十章 生态龕重叠指数

在一个群落中共存种如何利用共有的资源?具有相同资源利用形式的种可以认为是高度“重叠”的;那些具不同利用形式的物种被认为是低重叠的。生态学家导出各种指数以测度种重叠程度,试图获得对群落结构的了解。在本章中我们将介绍 Petraitis(1979)的重叠测度。

10.1 基本方法

从 Gause(1934)的经典工作开始,群落生态学家们长期以来一直在寻找了解共存种如何利用象食物或空间这样的共同资源的方法。这样的研究导出了关于群落中种多样性自然规律的理论,并努力改善物种生态龕的概念(Schoener 1974)。回忆第七章,将物种生态龕定义为:物种在群落中的“位置”,例如包括对食物资源的利用,活动时间,地点,与其他种相互作用的方式,等。

早期试图证实物种依生态龕而共存这一要点的人之一是 Rianka(1973)。Pianka 要测度各种被比较的共存蜥蜴种与选择食物的大小以及觅食地点和时间有何关系。他引入了一个经验指数以测度与这些资源状态有关的不同种的相似性或重叠的程度。原则上,具有相似生态龕需求的两个种可以期望表现出高度重叠。在理论上把龕重叠认为是物种多样性和群落结构的可能决定因素之一(petraitis 1979)。

龕重叠指数现为生态学家广泛应用,无疑文献中有各种各样的建议形式。在第九章中我们把重叠指数划分为 R-型方法。但传统中用于 Q-型研究的指数也被用于测度重叠。就广意而言,这些重叠指数可被划分为距离测度(如 Levins 1968)、关联指数(如 Cody 1974)、相关系数(Pianka 1973)、信息测度(如 Horn 1966)和检验统计(如 van Bell 和 Ahmad 1974)见图 9.1)。关于这些方面的论述可以在 Pielou(1972a),Abrams(1980),Hurlbert(1978),Lawlor(1980),Zaret 和 Smith(1984)的文献中找到。

Hurlbert(1978)曾建议:重叠指数的选择应具有容易进行生物学解释和对资源状态可用性的变化进行解释的能力。考虑到后者,例如可以根据种的形态、觅食时间、消耗的食物资源、利用的微生物等(Lawlor 1980)[这些被定义为资源状态(resource states)]进行比较。我们在表 10.1 中举出例子,其中把四个目的昆虫作为鸟种的(即食肉动物)资源状态(即被捕食动物)。可以看出鸟种 1 均匀地取食这些被捕食的动物,而种 2 和种 3 专门地分别觅食被捕食动物 1-2 和 3-4。但是,当考虑到资源状态的相对多度时,则出现更有信息量的解释。种 1 拒绝共同的资源项目,即捕食项目 1 和 2(占总资源库的 80%),就这一点而言,种 1 同等地利用全部资源。鸟种 2 专门食用丰富的被捕食动物项 1 和 2,所以当这个种随机地遇到环境中的潜在被捕食动物时仅排斥其中的五分之一。鸟种 3 将排斥五分之四,因为它宁愿选择最不丰富的被捕食动物。当考虑到资源的可利用性时,学生们应能直观地看到:鸟种 1 与 2 利用资源的“重叠”和 1 与 3 的是非常不同的。

表 10.1 3 个鸟种捕获 4 种资源状态(昆虫目)的假设数据, 改自 Petraitis (1979)

昆虫目		被捕食的昆虫项(资源状态)			
		(1)	(2)	(3)	(4)
相对可利用的资源		0.4	0.4	0.1	0.1
		种			
鸟种的相对利用量	(1)	0.25	0.25	0.25	0.25
	(2)	0.50	0.50	0	0
	(3)	0	0	0.50	0.50

Schoener (1974)和 Hurlbert (1978)提出用资源的可用性为每一物种对资源的相对利用量加权。然而大多数重叠指数都是以对资源状态相对利用量为依据,而不考虑相对可用性(即通常假定资源是同等可用的)。因而,必须小心地使用这些系数(Hurlbert 1978, Lawlor 1980, Petraitis 1979)。当然必须意识到:这类可用性数据在野外研究中是很少能得到的。进而,我们主观确定的一个物种的“相对”可用性可能或不可能反映特定物种的实际利用情况。认识到这些限度,根据一个种利用资源与其他种相等的可能性, Petraitis (1979)发展了对重叠的测度。下面讨论他的特定重叠(specific overlap)和普遍重叠(general overlap)。

10.2 步骤

每个种对资源状态的相对利用量被称为利用曲线(utilization curve)。例如,在表 10.1 中,鸟种 3 的利用曲线是[0,0,0.5,0.5],种 1 是[0.25,0.25,0.25,0.25]。

Levins (1968)提出测度利用曲线重叠程度的指数。种 1 和种 2 重叠的 Levins 指数(LO)为:

$$LO_{1,2} = \frac{\sum_j [(P_{1j})(P_{2j})]}{\sum_j (P_{1j}^2)} \tag{10.1}$$

其中术语的定义在表 10.2 中。就我们在表 10.1 中所举例子而论

$$LO_{1,2} = \frac{(0.25)(0.50) + \dots + (0.25)(0)}{0.25^2 + \dots + 0.25^2} = 1.0$$

同样种 2 与种 1 的利用曲线的重叠是:

$$LO_{2,1} = \frac{(0.25)(0.50) + \dots + (0.25)(0)}{0.25^2 + \dots + 0^2} = 0.5$$

所以种 1 的利用曲线完全与种 2 的重叠,但种 2 的利用曲线只有一半与种 1 的相重叠。这种差异是由如何使两个种之间重叠的量[即方程 (10.1) 的分子]标准化所致,方程 (10.1) 中的分母被称为第 i 种的宽度 (breadth) (B) 为:

$$B_i = \frac{1}{\sum_j (P_{ij}^2)} \tag{10.2}$$

注意:种 1 的宽度(见表 10.1)是 4.0,而种 2 的宽度是 2.0。还有,方程 (10.2) 可能显得

很熟悉，如同在第 8 章中出现的相似 [方程 (8.5c) 和 (8.6)]。

表 10.2 计算特定的和普遍的重叠测度的术语总结。3 个资源组($r=3$)为两个种($S=2$)所利用。下标 i, j 代表第 i 种和第 j 资源修改于 Petraitis(1985)

种	资源组(r)			总计
	(1)	(2)	(3)	
计数 (1)	$n_{1,1}$	$n_{1,2}$	$n_{1,3}$	N_1
(2)	$n_{2,1}$	$n_{2,2}$	$n_{2,3}$	N_2
总计数	t_1	t_2	t_3	T
比例 (1)	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	
(2)	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	
组合比例	c_1	c_2	c_3	

$$N_i = \sum_{j=1}^r (n_{ij}) \quad t_j = \sum_{i=1}^S (n_{ij}) \quad T = \sum_{j=1}^r (t_j)$$

$$p_{ij} = n_{ij}/N_i \quad c_j = t_j/T$$

$$B_i = \frac{1}{\sum_j (P_{ij}^2)} \tag{10.2}$$

Levins 指数测度的重叠不考虑资源可用性的差异。Hurlbert (1978) 试图通过用资源的可用性给重叠加权的方法来使之得以纠正。种 1 和种 2 间的重叠应用 Hurlbert 指数 (HO) 即为：

$$HO_{1,2} = \sum_j \frac{(P_{1j})(P_{2j})}{C_j} \tag{10.3}$$

其中 C_j 是第 j 资源的相对多度 (见表 10.2)。利用表 10.1 的数据，由 HO 再次测度的重叠是

$$HO_{1,2} = \frac{(0.25)(0.50)}{0.4} + \dots + \frac{(0.25)(0)}{0.1} = 0.625$$

种 2 利用曲线与种 1 的重叠是

$$HO_{2,1} = \frac{(0.50)(0.25)}{0.4} + \dots + \frac{(0)(0.25)}{0.1} = 0.625$$

在此情况下，重叠为资源的可用性而不是象 Levins 指数那样为种宽度所标准化。所以 $HO_{1,2} = HO_{2,1}$ 。Hurlbert(1982)指出：在“两个种按其多度的比例利用每一资源状态”的情况下，HO 将等于 1。

10.2.1 特定重叠

象前面的 LO 和 HO 一样，特定重叠(SO)也以利用曲线的比较为根据。但又与 LO 和 HO 不同，种 i 特定重叠于种 k 之上的数量是种 i 的利用曲线能够从种 k 的利用曲线中画出的概率。petraitis 通过检查被观察种对资源状态的利用是否可以从环境资源谱中随机地画出而导出下面的方程。利用在表 10.2 中阐明的那类种-资源数据，有可能计算出物种 i 利用的(即

n_{i1}, n_{i2}, n_{i3})资源可以从物种 k 的利用曲线中画出的概率(即 P_{k1}, P_{k2}, P_{k3})。

步骤 1. 提出假设。 可以检验两个一般假设类型:(1)两个种完全重叠的零假设,即它们的利用曲线相等;(2)种 i 在一个种(k)之上的特定重叠大于种 i 在另一个种(m)上的重叠的零假设。

步骤 2. 计算特定重叠。 为了便于介绍,特定重叠由一个种对(种 1 和种 2)示出。全部术语及其定义在表 10.2 中列出。种 1 在种 2 之上和种 2 在种 1 之上与 r 个资源组有关的特定重叠分别给定为

$$SO_{1,2} = e^{E_{1,2}} \quad (10.4)$$

和
$$SO_{2,1} = e^{E_{2,1}} \quad (10.5)$$

其中
$$E_{1,2} = \sum_j (P_{1j} \ln P_{2j}) - \sum_j (P_{1j} \ln P_{1j}) \quad (10.6)$$

和
$$E_{2,1} = \sum_j (P_{2j} \ln P_{1j}) - \sum_j (P_{2j} \ln P_{2j}) \quad (10.7)$$

注意:方程(10.6)和(10.7)的右端分量是香农函数[见方程(8.8)]。还有,要弄清计算 SO 需要两个种利用全部资源组,因为,如果一个种对一种资源的利用为 0(即 $P_{ij}=0$),则方程(10.6)或(10.7)中的 $\ln(P_{ij})$ 则无意义。最后 SO 的取值范围是 0 到 1。

步骤 3. 计算检验统计。 首先为了检验种 i 在种 k 上的特定重叠是完全的这一零假设,我们计算统计量

$$U_{i,k} = -2N_i \ln(SO_{i,k}) \quad (10.8)$$

其中 $U_{i,k}$ 是具 $r-1$ 自由度的卡方分布。所以对于种 1 在种 2 上的特定重叠,我们有

$$U_{1,2} = -2N_1 \ln(SO_{1,2})$$

而对于种 2 在种 1 上的则有

$$U_{2,1} = -2N_2 \ln(SO_{2,1})$$

其次,为了检验种 i 在种 k 上的特定重叠大于种 i 在种 m 上的重叠,我们按下式计算对数似然比率 W :

$$W = N_i \ln(SO_{i,k}/SO_{i,m}) \quad (10.9)$$

如果 $W > 2$,我们就可得出结论说种 i 在种 k 上的特定重叠大于 i 在种 m 上的重叠。

10.2.2 普遍重叠

群落中种间普遍重叠的量, Petraitis 定义为:全部种的利用曲线均从“共同”的利用曲线中画出的概率。被比较的种数未被限定(回忆:特定重叠是一个种的利用曲线可以从另一种的曲线中画出的概率)。种间普遍重叠按如下种利用曲线的加权平均计算:

步骤 1. 提出假设。 被检验的假设是种的重叠是完全的(即种的利用曲线可以由共同的利用曲线中画出)。

步骤 2. 计算普遍重叠(GO)。 普遍重叠由下式给出

$$GO = e^E \quad (10.10)$$

其中
$$E = \frac{\sum_i \sum_j [n_{ij} (\ln c_j - \ln P_{ij})]}{T} \quad (10.11)$$

(仍参考表 10.2 术语的定义)。注意:方程(10.11)中的连加贯穿了所有种($i=1$ 到 S)的所有资源组($j=1$ 到 r)。

步骤 3. 计算检验统计量。 统计量

$$V = - 2T \ln G O \tag{10.12}$$

是按具 $(S-1)(r-1)$ 自由度的卡方变量分布的。如果 V 超出卡方的临界值 $P=0.05$, 则完全重叠的零假设被拒绝。

10.3 例子：土蜂共同生存的计算

在科罗拉多土蜂 (*Bombus* spp.) 种间生态龕分化的研究中, Pyke (1982) 测定了蜂喙的长度与调查到的花冠长度之间的相互关系。在此例中, 用在四组花冠长度表示的资源谱上的四种土蜂的观察数进行计算 (表 10.3)。

表 10.3 利用四个资源组 (所调查花的花冠长度毫米数) 的四种土蜂 (*Bombus* spp.) 数据和特定重叠

计数 种		资源组(r)				总计
		(1)	(2)	(3)	(4)	
		(0-4mm)	(4-8mm)	(8-12mm)	(>12mm)	
B. appositus	(1)	27	47	357	925	1,365
B. flavifrons	(2)	1,018	1,363	1,139	964	4,484
B. frigidus	(3)	333	638	145	13	1,129
B. occidentalis	(4)	155	84	70	51	360
总计		1,533	2,132	1,711	1,953	7,329
比值						
B. appositus	(1)	0.020	0.035	0.263	0.682	
B. flavifrons	(2)	0.227	0.304	0.254	0.215	
B. frigidus	(3)	0.295	0.565	0.128	0.142	
B. occidentalis	(4)	0.431	0.233	0.194	0.142	
组合比值		0.209	0.291	0.233	0.266	
特定重叠(SO_{ijk})	种	(1)	(2)	(3)	(4)	
	(1)	1.0	0.510	0.059	0.359	
	(2)	0.385	1.0	0.574	0.912	
	(3)	0.107	0.736	1.0	0.736	
	(4)	0.226	0.902	0.675	1.0	

10.3.1 计算特定重叠

步骤 1. 提出假设。 我们可以检验这样的假设,即假设任何蜂种对间的特定重叠是“完全的”——相对于可进行选择的“有些”或“非”。进而,我们可以检验蜂种 i 在种 k 上的特定重叠大于种 i 在种 m 上的这样一种假设。在此例范围内重叠涉及土蜂对以花大小表示的资源组的选择。

步骤 2. 计算特定重叠(SO)。 为了详细说明,计算出蜂种 1 和 2 与 2 和 1 间的特定重叠。根据表 10.3 和方程(10.6),

$$\begin{aligned} E_{1,2} &= [(0.02 \ln 0.227) + \dots + (0.682 \ln 0.215)] \\ &\quad - [(0.02 \ln 0.02) + \dots + (0.682 \ln 0.682)] \\ &= -1.48 - (-0.81) = -0.673 \end{aligned}$$

根据方程(10.7)

$$\begin{aligned} E_{2,1} &= [(0.227 \ln 0.02) + \dots + (0.215 \ln 0.682)] \\ &\quad - [(0.227 \ln 0.227) + \dots + (0.215 \ln 0.215)] \\ &= -2.33 - (-1.38) = -0.955 \end{aligned}$$

因此,根据方程(10.4)

$$SO_{1,2} = e^{-0.673} = 0.510$$

并根据方程(10.5)

$$SO_{2,1} = e^{-0.955} = 0.385$$

利用 Levins 指数对应值是 $LO_{1,2}=0.427$ 和 $LO_{2,1}=0.898$ 。当用 LO 进行测度时, *B. appositus* 在 *B. flavifors* 之上的以及反过来的这些非常不同的重叠是由这两个种差异很大的龕宽度所导致的 ($B_1=1.87$ 和 $B_2=3.13$)。另一方面,这两个种的 SO 值是非常相似的,通常情况都是这样,因为没有应用这样的标准化量。这些数据的其他特定重叠在表 10.3 中给出。

步骤 3. 计算检验统计量。 按照方程 (10.8)

$$U_{1,2} = (-2) \times 1356 \times (-0.673) = 1826$$

和 $U_{2,1} = (-2) \times 4484 \times (-0.955) = 8564$

二者均远高于临界值 7.82 ($df=3, P=0.05$), 所以, 完全重叠的假设被拒绝。这是或多或少受到限制的检验, 因为选择要么是“非”重叠, 要么是“有些”重叠 (即: 相对重叠强度仅针对“完全”重叠而被检验)。

Pyke (1982) 测定了蜂种的平均喙长度 (单位 mm), 以观察是否在被调查的喙长与花的花冠长度之间存在相关关系。在表 10.3 中四个蜂种的喙长度可以按从最大到最小排成序列 $1 > 2 > 3 = 4$ 。根据不同花冠长度的花的调查数, 我们可以期望不同喙长度的土蜂间的龕分离。利用对数似然比率 W [方程 (10.9)], 我们可以检验是否一个种在另一种上的重叠大于这个种在其他种上的重叠。例如, 我们可以检验以下两个假设:

情况 (1) ——种 1 (最大喙长) 在种 2 (中等喙长) 上的重叠大于种 1 在种 3 (最小喙长) 上的重叠。对数似然比率 $W=1356 \ln(0.51/0.059)=2925$, 远大于临界值 2。所以我们接受这一假设。

情况 (2) ——种 2 (中等喙长) 在种 3 (最小喙长) 上的重叠大于种 2 与种 4 (也是最小

的)的重叠。对于似然比率 $W = 418.4 \ln(0.574/0.912) = 2076$ ，故拒绝这种假设。

10.3.2 计算普遍重叠

步骤 1. 提出假设

被检验的假设是在四个蜂种对花冠长度进行选择上完全重叠。根据特定重叠的计算结果，我们期望拒绝这一假设。

步骤 2. 计算普遍重叠。 利用方程 (10.11)

$$\begin{aligned} E &= [27(\ln 0.209) - \ln 0.20] + 47(\ln 0.291 - \\ &\quad \ln 0.035 + \dots + 51(\ln 0.291 - \ln 0.142)]/7329 \\ &= -1240.3/7329 = -0.169 \end{aligned}$$

因此,按照方程(10.10)获得 GO(记作GO)的估计值是

$$GO = e^{-0.169} = 0.844$$

步骤 3. 计算检验统计量。 利用方程 (10.12), $V = (-2)(7,329)(-0.169) = 2,480$ 的值远大于临界卡方值 16.92 ($df=9$, $P=0.05$)。所以,虽然重叠指数 ($GO=0.84$) 很大,我们仍拒绝完全重叠的假设。正象先前所述,针对“完全”重叠检验蜂种间的相对重叠强度。

10.4 例子: 鸟食谱重叠

Root (1967) 检查了三种鸟的胃肠内容物以检查它们在捕食选择中的相似性 (表 10.4)。这个例子被 Petraitis (1979) 用来说明特定重叠及普遍重叠的测度。

我们的 BASIC 程序 SPOVRLAP. BAS (见附磁盘) 的输出内容在表 10.5 中给出。三个鸟种间在食性选择上的普遍重叠很高 ($GO=0.85$), 但远非“完全”, 因为 $V=211.2$, 而卡方的临界值是 15.5 ($df=8$, $P=0.05$)。 *Vireo huttoni* (种 3) 在 *Poliptila caerulea* (种 1) 上的特定重叠是 0.707, 而它在 *Vireo gilvus* (种 2) 上的重叠是 0.711。利用对数似然比率 $W = 134 \ln(0.707/0.711) = 10.16$, 我们可以得出结论说: 在食性选择上 *Vireo huttoni* 表现出与 *Poliptila caerulea* 的重叠大于与 *Vireo gilvus* 的重叠。

表 10.4 三个共存鸟种食谱中的节肢动物。数据引自 Root (1967)

种	资源 (被捕食的昆虫)				
	(1) 半翅目	(2) 鞘翅目	(3) 鳞翅目	(4) 膜翅目	(5) 其他
(1)	103	93	20	40	31
(2)	23	31	132	14	13
(3)	16	40	33	30	15

* (1) = *Poliptila caerulea*
(2) = *Vireo gilvus*
(3) = *Vireo huttoni*

表 10. 5 利用 BASIC 程序 SPOVRLAP. BAS 计算的普遍重叠和特定重叠，
引自 Root (1967) 的鸟类数据

A. 种的普遍重叠					
种数	GO	G _{min}	G _{adj}	V	df
3	0. 847	0. 349	0. 764	211. 2	8
B. 特定 (成对的) 重叠					
种	对	指数和检验统计量			
i	k	SO	U		df
1	2	0. 494	404. 4		4
2	1	0. 359	436. 1		4
1	3	0. 768	151. 3		4
3	1	0. 767	71. 0		4
2	3	0. 713	144. 4		4
3	2	0. 711	91. 5		4

10. 5 重叠问题的讨论

为了阐述普遍重叠行为的一些有趣方面，我们在表 10. 6 中列出了四种不同的数据组。

表 10. 6 两个种利用两种资源的数据和特定的 (SO) 普遍 (GO) 最小的 GO (GO_{min})
和调整的 GO (GO_{adj}) 重叠指数

情况	种	资源组			指数			
		(1)	(2)	N _i	SO	GO	GO _{min}	GO _{adj}
A	(1)	10	0	10	0	0. 82	0. 82	0. 0
	(2)	0	190	190				
B	(1)	100	0	100	0	0. 52	0. 52	0. 0
	(2)	0	100	100				
C	(1)	50	0	50	0	0. 57	0. 57	0. 0
	(2)	0	150	150				
D	(1)	50	50	100	1	0. 50	0. 50	1. 0
	(2)	50	50	100				

在 A、B 和 C 情况下，两个种非共享共同的资源，而在 D 情况下它们的利用曲线是一样的。由于本章的许多方程不允许使用 0 值，我们把 0 变成 1×10^{-7} ，并利用 SPOVRLAP. BAS 程序计算 GO 和 SO 值。对于情况 A—C，种 1 在种 2 (和反之) 上的特定重叠为 0，对于情况 D，正如期望一样特定重叠为 1。然而，尽管事实上在情况 A—C 下种 1 和种 2 之间在资源利用中没有任何重叠，普遍重叠 (GO) 从高 0.82 变化到低的 0.50 (表 10.6)。Smith (1984) 指出：

GO 的下限取决于样本含量和所考虑的种数。并证明根据 N_i 和 T (Smith 1984) 可以算出 GO 的最小值, 为:

$$GO_{min} = e^{(1/T)[\sum_{i=1}^S (N_i \ln N_i) - T \ln T]} \quad (10.13)$$

所以, 在情况 A 中 GO_{min} 是 0.82, 而在情况 B 为 0.50。Smith (1984) 介绍说考虑到对样本大小的依赖。应调整普遍重叠 GO 的估值。一个可能性是调整与 GO_{min} 的可能影响范围有关的 GO 值, 为

$$GO_{adj} = \frac{GO - GO_{min}}{1 - GO_{min}} \quad (10.14)$$

就表 10.6 所举例子而言, 在无重叠 (A-C) 情况下 GO_{adj} 的值为 0, 在完全重叠 (D) 情况下, GO_{min} 是 1。 GO_{adj} 的值总限于 0 到 1, 而在两物种情况下 GO 将从 0.5 到 1 之间任意变化。在两种以上的情况下, GO 的低限将下降 (Smith 1984)。但是注意: 对于统计检验来说, GO 的未调整值总被应用 (Petraitis 1985)。

在阅读生态学文献时学生们将遇到许多重叠系数, 通常以最初作者命名 (如 Levins, Pianka, Morisita, Horn, 等)。大多数重叠测度的限制是难以确定重叠程度在统计学上是否显著 (Zaret and Smith 1984)。经常提出的问题是: 在何值处有一个有效地与 0 重叠 (无重叠) 分离的重叠指数以指出重叠的显著程度? 由 Petraitis (1979, 1985) 发展的特定的和普遍的重叠指数是解决这一问题的方法之一。Garratt 和 Steinhorst (1976) 根据非参量置换检验的统计方法描述了一种旨在检验 Moissita 指数的显著性和置信区间的技术。Zaret 和 Smith 提出了计算检验统计量和包括 GO 在内的几个重叠测度置信区间的分析方法。最后, Mueller 和 Altenberg (1985) 利用刀切 (Jackknife) 法和自展法 (bootstrap) 估计 Moissita 和 Horn 重叠指数中方差和误差。

10.6 总结和推荐

1. 重叠指数为生态学家所广泛使用。但是大多数重叠指数的有用性是有限的, 因为这些指数不考虑资源的可用性 (全部资源被假定为同样可用), 也没有可用的统计检验 (第 10.1 节)。

2. 至于资源的可用性, 在野外研究中很少能获得这样的数据。进而, 我们观察到的“可以利用”或者可能, 或者不可能反映物种的利益所在。因为生态学研究总是实事求是的, 至于什么是“重要”数据, 一定要通过对我们如何为最初的感知力所限制这一问题进行深思熟虑来完成对数据的解释和计算。

3. Petraitis (1979) 引进了特定重叠 (SO) 的测度, 这是以一个种的利用曲线 (即利用每一资源的比例) 可以从另一物种的资源曲线上画出的概率为基础的 (10.2.1 节)。

4. 对于检验一个种在另一种上特定重叠是完全的这一零假设, 统计检验是可用的。但是, 如果我们拒绝这一假设, 则选择只能是“无”或“有些”, 稍微限制了这种检验的有用性 [方程 (10.8)]。

5. 对数似然比率是使我们能够检验一个种在某一种上的重叠大于这个种在另一个种上的重叠这一假设的有用统计量 [方程 (10.9)]。

6. 由 Petraitis (1976) 发展的普遍重叠测度值 GO 是以两个或更多的种的利用曲线自“共同的”利用曲线中画出的似然率为根据的 (10.2.2 节)。

7. GO 的最小值可以假定是样本含量和种数的函数。所以, 我们介绍了考虑到特定数据组的可能范围调整的 GO 。在方程 (10.14) 中我们提供了调整的可能形式, 从而 GO_{adj} 的取值是 0 到 1。

8. 由于在计算中用到对数, 故无论是重叠的特定测度 (SO), 还是普遍测度 (GO), 都要求全部物种利用全部资源 (即数据中无零)。严格地讲, 由于在生态学数据组中存在零是潜在的必然, 因此限制了这些指数的有用性。

9. 一定要小心地解释重叠指数, 尤其是涉及到竞争强度时。正如 Lawlor (1980) 与 Zaret 和 Rand (1971) 指出的: 在低重叠的一些测度间, 即竞争强度, 可能存在着不可解释的相关, 例如, 如果这种低重叠是“过去”竞争的结果。我们建议学生们读一读 Lawlor (1980) 的文章。

第十一章 种间关联

种的相互作用在物种的生态学中占有重要的地位。在任何给定的群落内，有许多影响种的分布，多度以及种间相互作用的生物的和非生物的因素。无论如何，依是否两个种选择或避免存在于同样的生境，具有某些相互吸引或排斥的性质，或不存在相互作用，都导致一定的种间关联 (interspecific association) 格局。这种关联可以是正的、负的或不存在的。在这一章中我们叙述检验种间存在关联的方法并提出测度关联程度的指数。这些技术完全依据于取样单位 (SU_s) 中物种的存在与否。

11.1 基本方法

在此我们涉及到测度两个种在相同位置被发现的频率是多少的问题。这种两物种共存的亲合性 (或缺乏亲合性) 被称为种间关联。通常，两个种间存在关联是因为：(1) 两个种都选择或都避免相同的生境或生境因子；(2) 它们对一般性非生物的和生物的环境具有相同的需求；(3) 一个种或两种都对其他种具有或者是吸引或者是排斥的亲合性 (Hubalek 1982)。

表 11.1 可以导致物种间正和负关联的生态过程和种间相互作用。摘自 Schluter (1984)

相互作用	过程举例	
	负的	正的
无	物种具有不同的资源需求	物种对非限制性资源的利用有共同的反应
互惠共生	资源竞争并为物种排它性地利用	物种互相促进提高生存概率
竞争	物种间妨碍，产生非经常的排斥	物种对有限资源的反应呈一致波动
捕食	高密度捕食者导致当地被捕食者衰败	捕食者对被捕食者变化的反应呈正的波动

查明种的关联具有重要的生态学涵义。在表 11.1 中总结了一些可以导致两物种间具正或负关联的生态学过程。特别要注意：即便在本身缺乏种间相互作用的情况下也可能存在正的或负的关联。还要注意：“无关联”的结果显然未包括在此清单中，因为这可能起因于正负作用力的平衡 (Schluter 1984)。如同本教材所强调的，格局 (例如种间联系) 的查明并未提供为什么这样的格局可能存在的原因。而理想的格局检查，应导致产生这种格局潜在的因子的假设的产生，可设计辅助研究以达到这一目的。

种的关联的研究包括两个不同的内容。首先是在预定的概率水平上对两个种有或无关联的假设进行统计检验。其次是关联程度或强度的测度 (图 9.1)。这些被认为是关联的不相关

联特征。下面，我们（1）概括用于检验种对间关联的步骤，（2）提出三个种间关联的测度，（3）说明在许多种中查明关联的联合检验。

11.2 步骤

研究种间关联的步骤是根据种在成组的 SU_s 中存在或不存在来进行的。我们可以用二进制数来表示，即存在记作 1，不存在记作 0。 SU_s 可以是自然的如羽毛、朽木、叶子，也可以是人为的如样方、样地、样线。（回忆在第二章中我们讨论了许多有关在空间格局分析中利用人为 SU_s 的问题；在有关关联的问题上也存在同样的问题）。依 SU 的大小和形状不同，可能会影响关联的结果。如选择与被研究种的大小形状和空间分布有关的 SU ，这种影响将会减小。 SU 必须足够大以潜在地包括每个种的一个个体，但又不能太大以至大到在每个 SU 中都包含着这些种中的一个（Greig-Smith 1983）。

11.2.1 关联的检验（二物种的情况）

步骤 1. 数据整理。对于每一种对 A 和 B，我们求得如下：

- a=两个种都出现的 SU_s 数；
- b=种 A 出现，种 B 不出现的 SU_s 数；
- c=种 B 出现，种 A 不出现的 SU_s 数；
- d=种 A 和 B 均未出现的 SU_s 数；
- $N=SU_s$ 的总数 ($N=a+b+c+d$)。

这种信息以 2×2 表（图 11.1）的形式整理是很方便的。下面示出的关联的检验和测度都以这些数据为基础。

种 A 在 SU_s 中出现的期望频率，我们以 $f(A)$ 表示，给定如下：

$$f(A) = \frac{a + b}{N} \tag{11.1a}$$

而种 B 为

$$f(B) = \frac{a + c}{N} \tag{11.1b}$$

我们假定在取样时两个种至少在一个 SU 中出现过，即 $f(A)$ 和 $f(B)$ 大于 0。

		种 B		
		出现	不出现	
种 A	出现	a	b	m=a+b
	不出现	c	d	n=c+d
		r=a+c	s=b+d	N=a+b+c+d

图 11.1 2×2 列联表或物种关联表

步骤 2. 提出假设。零假设为种是独立的（即无联系）。

步骤 3. 计算检验统计量。 2×2 列联表含有来自样本容量为 N 的每一格子 (a, b, c 和 d) 中的观察值。为了检验是否存在关联, 我们计算每一格中将可能出现的期望值——如果种 A 和 B 的出现是独立的并把它所与观测值进行比较。卡方检验统计量可被用来检验 2×2 列联表中独立性的零假设。卡方检验统计量按下式计算:

$$\chi^2 = \sum \frac{(\text{观测值} - \text{期望值})^2}{\text{期望值}} \quad (11.2)$$

这是 2×2 列联表各格子的总和。

a 格的期望值给定为

$$E(a) = \frac{(a+b)(a+c)}{N} = \frac{rm}{N} \quad (11.3)$$

或根据方程 (11.1)

$$E(a) = f(B)(a+b) = f(A)(a+c) \quad (11.4)$$

简言之, 方程 (11.4) 表明种 A 存在于 SU_s 中的总数 (即 $a+b$)。我们期望: 如果 A 和 B 是独立的, 种 B 也应按比例在 SU_s 中以其总频率出现, 即 $f(B)$; 而种 A 在种 B 出现的 SU_s 中的出现也是如此。

同理, 格 b, c 和 d 的期望值分别为

$$E(b) = \frac{ms}{N}, E(c) = \frac{rn}{N}, E(d) = \frac{sn}{N} \quad (11.5)$$

卡方检验统计量 [方程 (11.2)] 现给出如下

$$\chi^2 = \frac{[a - E(a)]^2}{E(a)} + \dots + \frac{[d - E(d)]^2}{E(d)} \quad (11.6)$$

一个在数学上等价的, 当然较为简单的方程 (可以代替方程 (11.6) 使用) 是

$$\chi^2 = \frac{N(ad - bc)^2}{mnr s} \quad (11.7)$$

实际上, 除较为简单便于使用外, 由于方程 (11.7) 不要求计算期望值, 也没有观测值和期望值的差异, 可避免关联的舍入误差。

比较卡方检验统计量与理论卡方分布来确定卡方检验统计量的显著性。由于具 r 行和 c 列的列联表具有 $(r-1)$ 乘 $(c-1)$ 自由度, 故 2×2 列联表的自由度为 1 (Zar 1974)。1 自由度在 5% 概率水平处的卡方理论值为 3.84。如果 $\chi^2 > 3.84$, 我们拒绝关于种 A 和 B 的共同出现是独立的这一零假设, 从而得出结论说它们是有关联的。

关联有两种类型:

1. 正的关联——如果观测的 $a > E(a)$, 即种对同时出现的次数比在独立情况下期望的次数更多。
2. 负的关联——如果观测的 $a < E(a)$, 即种对同时出现的次数比在独立情况下期望的次数少。

观测的 a 与 $E(a)$ 的这种比较, 即

$$a - E(a) = (ad - bc)/N \quad (11.8)$$

导出出现在如方程 (11.7) 那样的所有 χ^2 型公式的分子中的量 $ad - bc$ 。

如果在 2×2 列联表的任一个格子中具有 < 1 的期望频率或如果两个以上的表格具有 < 5

的期望频率, 则导出的卡方检验统计量将是有偏的 (Zar 1974)。一种校正卡方被用来避免由低格内的期望值导致的偏差。在这种情况下用连续性校正来保证较紧密地逼近理论的, 连续的卡方分布。通过应用 Yates 的校正公式可以达到这一目的:

$$X^2_i = \frac{N[|(ad) - (bc)| - (N/2)]^2}{mnrS} \quad (11.9)$$

11.2.2 关联的测度 (二物种的情况)

Hubalek (1982) 评论了 43 个被用来测度种对间关联程度的指数的特点。为了整理这些过剩的指数, Hubalek 确定了五项“允许”条件。不能满足任何一项条件的指数均被认为是不允许的且不再作进一步考虑。然后将剩余允许的指数与八项任选准则进行比较, 以便有助于选出最好的关联指数。Janson 和 Vegelius (1981) 进行了同样的研究, 研究中 20 个关联指数的特性由六项“允许”条件检验。全部这些允许条件的细节超出了本书的范围, 但五项重要的条件在此列出:

- 条件 1. 在 $a=0$, 即当两个种决不同时出现时, 每个关联指数应达到最小值。
- 条件 2. 当两个种总是同时出现时, 即 $b=c=0$ 时指数应达到最大值。
- 条件 3. 关联指数应是对称的, 即不论将哪个种设为“ A ”或“ B ”, 指数的值都应相同 (图 11.1)。
- 条件 4. 指数应能够区分出正的和负的联系。形式上这意味着当 $a > E(a)$ 时的指数值总是大于当 $a < E(a)$ 时的。

条件 5. 指数应独立于 d (即同时不出现)。关于种的同时不存在是否有着生态学含义问题曾有许多争议 (Clifford 和 Stephenson 1975, Goodall 1978b, Sneath 和 Sokal 1973)。我们与 Hubalek (1982) 在生态学中对使用 d 值的指数加以限制这一观点一致。例如, 在由 Bultman 和 Faeth (1985) 所做的栎树叶上潜叶蝇的研究中, 用来检验关联的 2×2 列联表的平均值是 $a=24$, $b=875$, $c=1140$, 而 $d=134650$, 任何使用 d 的指数都会被 d 巨大数量所淹没。

Hubalek (1982) 发现了满足他的允许条件的六种关联测度, Janson 和 Vegelius (1981) 发现了通常表现良好的三种。两项研究所介绍的三种测度——Ochiai, Dice 和 Jaccard 指数——在下面谈到。这些指数在“无关联”时等于 0, 在“最大关联”时为 1。Ochiai 和 Dice 测度是比值 a/m 和 a/r 的平均值, 即两个种同时出现的数与种 A 和种 B 分别出现的总数比较 (图 11.1)。

Ochiai 指数 (OI)。Ochiai (1957) 指数以 a/m 和 a/r 的几何平均值为基础, 即

$$OI = \frac{a}{\sqrt{a+b} \sqrt{a+c}} \quad (11.10)$$

Dice 指数 (DI)。Dice (1945) 指数以 a/m 和 a/r 的调和平均值为基础, 即

$$DI = \frac{2a}{2a+b+c} \quad (11.11)$$

Jaccard 指数 (JI)。这个指数是两个种都出现的 SU_s 数与至少出现一个种的 SU_s 总数之比:

$$JI = \frac{a}{a+b+c} \quad (11.12)$$

为了确定大量关联测度的取样特点, Goodall (1973) 采取从已知物种频度 (a, b, c 和 d) 的总

体重复取样的方法并计算每个指数的平均值和方差。发现即使在样本容量小($N=10$)的情况下, Jaccard 指数通常是无偏的。在小样本情况下 Dice 指数倾向低估实际种群值, 但在 $N=20$ 时表现良好。Goodall 不曾检验 Ochiai 指数。

11.2.3 种间关联(多物种情况)

通常不只单一种对之间的关联是有趣的; 我们可能对 5 个, 也许 50 个或更多的种感兴趣。全部种间关联的可能种对数或可以计算的组合数会按 $S(S-1)/2$ 方程迅速地增加, 其中 S 是种数。例如, 5 个种的组合数有 $5(4)/2=10$; 10 个种则为 $10(9)/2=45$ 。显然, 就易于说明而论, 这样的方法中存在着描述全部种对联系指数值的问题。多物种联系有两种图解法。

图解 1. 物种关联比较矩阵。以图 11.2 的形式表示的矩阵可以展示出全部种间关联的组合对。为了有助于解释, 种在矩阵中的位置可以按这样的方法重新排列, 即把具非常显著的正指数的种沿矩阵的对角线放置。

图解 2. 物种星座(网状)图 (Species constellation Diagrams)。二维图或网状图 (Plexus diagram) (McIntosh 1978) 可用来总结在群落中发现的成对种间的关联。网状图是这样构成的, 即那些具正关联的种紧凑地放在一块, 而那些具负关联的则远远地放在另一处。例如, 如果种 A 和 C 同 B, D 和 E 一样具正的关联, 且如果种 C 与种 D 具弱的关联, 则网状图如图 11.3 所示。

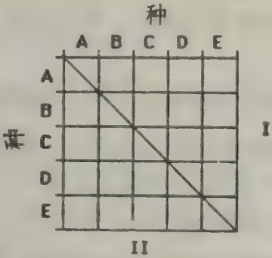


图 11.2 表示全部成对的种间关联矩阵。如算出两个指数, 则把其中一种置右上(I)并把另一种置左下(II)中是很方便的

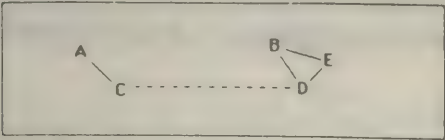


图 11.3 网状图示五个种(A-E)间假设的关联

种间距离可按关联指数的大小按比例绘制。例如, 种间距离的尺度可取反映 OI, DI 和 JI 值的 0 和 1 之间的数按比例绘图。由于这些图是用试错法画出的, 随着种数增加, 故需付出更大的努力来减少扭曲。

11.2.1 节中描述的关联检验适用于单一种对。但是, 当 $S>2$ 时这种成对技术就不适用

了。虽然我们可以计算种间关联的全部成对组合,但它们不是独立的,继而,我们不能指定结果分布的概率(Schluter 1984)。Pielou(1974)为这一目的讨论了过于苛刻的卡方值的使用。但是,检验许多成对种间关联的显著性的选择是考虑许多同时取得的 S 种间关联的显著性。Pielou (1972b)提出使用“2 到第 S”联表,但随着种数的增加这种方法变得不适用了。Schluter(1984)提出一种新方法,即使用由零关联模型(null association model)导出方差比率(VR)来同时检验显著的关联。关联的 VR 指数易于由存在一不存在数据导出,Schluter(1984)又证明了与无关联的期望值明显分离的检验统计量(W),其中 W 近似卡方分布。Schluter 的 VR 检验叙述如下:

步骤 1. 数据整理。 在表 11.2 中给出例子表示在 N 个 SU_s 中 S 种存在或不存在的数据库矩阵。

表 11.2 在 N 个 SU_s (j=1,2,3,...,N) 中 S 种 (j=1,2,3,...,S) 的存在(1)或不存在(0)的例子

种	SU _s					种总和
	(1)	(2)	(3)	...	(N)	
(1)	1	0	1		0	n ₁
(2)	1	0	1		1	n ₂
(3)	0	1	0		0	n ₃
⋮	⋮	⋮	⋮		⋮	⋮
(S)	0	0	1		1	n _s
SU 总和	T ₁	T ₂	T ₃	...	T _N	

步骤 2. 提出假设。 零假设是 S 种间无关联。在下面两个条件下这一假设为真:(1)种是独立的,(2)种间的正负关联抵消(Schluter 1984)。所以选择假设是种间存在“净”正或负联系。

步骤 3. 计算检验统计量 首先我们按下式计算 S 种在样本中出现的总样本方差:

$$\sigma_T^2 = \sum_{i=1}^S P_i(1 - P_i) \tag{11.13}$$

其中 P_i=n_i/N (参考表 11.2 术语的定义)。其次我们按下式估计总种数的方差:

$$S_T^2 = \frac{1}{N} \sum_{j=1}^N (T_j - t)^2 \tag{11.14}$$

其中 t 是每样本中种的平均数。

方差比率,

$$VR = S_T^2 / \sigma_T^2 \tag{11.15}$$

充当全部种的关联指数。在独立性零假设条件下期望值为 1。VR>1 表明种间表现出正的关联。如果 VR<1 表明存在负的净关联。

计算可用来检验偏离 1 是否显著的统计量 W。例如,如果种不相关联,则 W 落入由下面卡方分布给出的界限内的概率有 90%:

$$\chi^2_{0.05,N} < W < \chi^2_{0.95,N} \tag{11.16a}$$

其中

$$W = (N)(VR) \quad (11.16b)$$

Schluter(1984)指出可能存在这种情况,即某些种在它们自己间以正的形式出现,但与其它种则为负的,而方差比率检验确定为无关联。当然,即使在一个大的种随机集体中,利用11.2.1节中的成对卡方过程也将有一些显著的关联被查出;所以方差比率检验适用于种的较大集合。还有,当成对的检验无能为力时,这种检验将常常查出显著的关联。

11.3 举例:计算

在表11.3中给出了从五个 SU_s 中取出的三个种的多度和出现—不出现数据。实际上决不会有人试图分析这样的小样本容量的数据矩阵,但为了通过种间关联的必要计算,达到提供指导的目的,这些数据是可作例证的。

表 11.3 在五个取样单位(SU_s)中三个种(Spp)多度(a)和出现—不出现(b)的生态数据矩阵

(a)多度							
Spp	SU_s					平均数	方差
	(1)	(2)	(3)	(4)	(5)		
(1)	2	5	5	3	0	3.0	4.5
(2)	0	3	4	2	1	2.0	2.5
(3)	2	0	1	0	2	1.0	1.0

(b)出现—不出现						
Spp	SU_s					总计
	(1)	(2)	(3)	(4)	(5)	
(1)	1	1	1	1	0	4
(2)	0	1	1	1	1	4
(3)	1	0	1	0	1	3
—	—	—	—	—	—	—
总和(T_j)	2	2	3	2	2	11

11.3.1 关联的检验 (二物种的情况)

步骤 1. 数据整理。种对 1 和 3 的 2×2 列联表是 (其他种也有一个表同样计算)

		种 3		
		出现	不出现	
种 1	出现	a=2	b=2	m=4
	不出现	c=1	d=0	n=1
		r=3	s=2	N=5

步骤 2. 提出假设: “种 1 和种 3 在五个 SU_s 中出现是独立的”。

步骤 3. 计算检验统计量。从方程 (11.3) 中, 我们注意到 $E(a) = (4)(3)/5 = 2.4$, 这

与观测值 $a=3$ 接近。按方程 (11.7),

$$X_1^2 = 5 \times [2 \times 0 - 2 \times 1]^2 / [4 \times 1 \times 3 \times 2] = 0.83$$

并用连续性校正[方程(11.9)]

$$X_1^2 = 5 \times [|2 \times 0 - 2 \times 1| - (5 \div 2)]^2 / [4 \times 1 \times 3 \times 2] = 0.052$$

注意: 连续性校正因子对卡方统计量的值有很大影响, 尤其是当样本容量小时 (如此处的情况)。由于卡方在 $df=1$ 时的临界值是 3.84, 虽然我们再次提醒学生们这个例子仅是举例, 但我们不拒绝零假设。

11.3.2 关联的测度

由于发现种 1 和种 3 是独立的 (无关联的), 则关联强度的测度是无意义的。但是为了举出计算例子的目的, 我们按方程 (11.10) - (11.12) 计算 Ochiai, Dice 和 Jaccard 指数:

$$OI_{1,3} = \frac{2}{\sqrt{2+2} \sqrt{2+1}} = 0.85$$

$$DI_{1,3} = \frac{2 \times 2}{2 \times 2 + 2 + 1} = 0.57$$

$$JI_{1,3} = \frac{2}{2 + 2 + 1} = 0.40$$

11.3.3 种间关联 (多物种的情况)

步骤 1. 整理表 11.3b 的数据

步骤 2. 假设三个种间无关联。

步骤 3. 首先按方程 (11.13) 计算样本中种出现的总样本方差:

$$\sigma_t^2 = (4/5)(1 - 4/5) + \dots + (3/5)(1 - 3/5) = 0.56$$

总种数 [方程 (11.14)] 中方差的估值是 [其中 $t = (2+2+3+2+2) / 5 = 2.2$]

$$s_t^2 = (1/5)[(2 - 2.2)^2 + \dots + (2 - 2.2)^2] = 0.16$$

所以, 方差比率是[方程(11.15)]

$$VR = \frac{0.16}{0.56} = 0.28$$

VR 提出种间存在净负关联。事实上, 就每个种对而言观测的 $a < \exp(a)$ 。为了检验这个对 1 的偏离, 我们 [按方程 (11.16b)] 计算

$$W = (5)(0.28) = 1.43$$

在假设无关联的条件下, W 落入界内的概率有 90% [方程 (11.16a)]

$$1.14 < W < 11.07$$

所以, 我们接受无关联的零假设。

11.4 例子: 巴拿马蟑螂

Wolda 等人 (1983) 用黑光灯作为取样单位诱捕在巴拿马共和国对来自以下六个地区的蜚蠊目 (Blattaria) 蟑螂种进行研究: 巴罗科罗拉多岛 (BCI)、拉坤伯瑞斯 (LC)、法吐那

(FORT)、包坎特 (BOQ)、麦拉玛 (MIR) 和考里恩特·格兰地 (CORG)。在表 11. 4b 中给出五个多食科 (Polyphagidae) 蟑螂种的出现—不出现数据。利用 BASIC 程序 SPASSOC. BAS (见附软盘), 算出所有种的成对组合的卡方和种间关联指数 (OI, DI 和 JI), 还有多物种间关联的方差比率检验 (表 11. 5)。

在无关联的零假设条件下, 方差比率是 1. 06, 基本上等于期望值 1. 0。检验统计量 $W = 6. 38$, 落入 1. 63—12. 6 的范围 [方程 11. 16)], 所以我们接受零假设。

表 11. 4 从六个巴拿马地区取得的五个蟑螂中的 (a) 多度或 (b) 出现—不出现的生态数据矩阵
(修改自 Wolda 等人, 1983)

(a)多度								
		取样地						
种	No.	BCI	LC	FORT	BOQ	MIR	CORG	平均
		(1)	(2)	(3)	(4)	(5)	(6)	
Ceuthobiella spp.	(1)	0	0	0	0	0	1	0.17
Compsodes cucullatus	(2)	14	38	1	1	4	0	9.67
Compsodes delicatulus	(3)	28	4	1	0	1	0	5.83
Buboblatta armata	(4)	7	0	0	0	0	0	1.17
Latindia dohrniana	(5)	68	29	0	0	11	24	22.00

(b)出现—不出现								
		取样地						
种	No.	BCI	LC	FORT	BOQ	MIR	CORG	
		(1)	(2)	(3)	(4)	(5)	(6)	
Ceuthobiella spp.	(1)	0	0	0	0	0	1	
Compsodes cucullatus	(2)	1	1	1	1	1	0	
Compsodes delicatulus	(3)	1	1	1	0	1	0	
Buboblatta armata	(4)	1	0	0	0	0	0	
Latindia dohrniana	(5)	1	1	0	0	1	1	

只有一个种对 (Ceuthobiella spp. 和 Compsodes cucullatus) 实际表现出卡方检验统计量大于 3. 84; 但是如同所期望的, 由于样本容量小这些值都有偏差 (注意 SPASSOC. BAS 程序检验偏差值) 如果我们仅用关联指数来寻找种的亲和性, 则可能冒轻易得出结论的风险, 例如, 种 2 和种 3 是高度相关的 (图 11. 4)。当然指数测度关联的程度, 但不象卡方那样提供关联的检验。这可能指明要用大样本进行附加研究来证明。

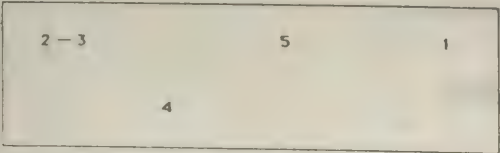


图 11. 4 来自巴拿马的五个蟑螂种的关联的网状图。由直线联结的种真正的关联

表 11.5 巴拿马六个地方上五种 Polyphagidae 蟑螂间的种间关联指数和检验统计量

多物种								
总体关联指数, $VR=1.06$								
检验统计量 $W=6.39$								
种对								
种对	关联类型 ^a	卡方		Yates	关联指数			
		值	误差	卡方	Ochiai	Dice	Jaccard	
1 2	—	6.00	*	0.96	0.00	0.00	0.00	
1 3	—	2.40	*	0.15	0.00	0.00	0.00	
1 4	—	0.24	*	0.96	0.00	0.00	0.00	
1 5	+	0.60	*	0.15	0.50	0.40	0.25	
2 3	+	2.40	*	0.15	0.89	0.89	0.80	
2 4	+	0.24	*	0.96	0.45	0.33	0.20	
2 5	—	0.60	*	0.15	0.67	0.67	0.50	
3 4	+	0.60	*	0.15	0.50	0.40	0.25	
3 5	+	0.38	*	0.09	0.75	0.75	0.60	
4 5	+	0.60	*	0.15	0.50	0.40	0.25	

a: 指示种间关联的方向

* 未校正的卡方值是有偏差的, 或者因为(1) 2×2 表的任何格子中的期望频率 < 1 和(或)(2)不止两个格子的期望频率 < 5

11.5 例子: 威斯康星森林

在这个例子中, 获自威斯康星南部十个森林小区的八个树种的出现—不出现数据被使用(表 11.6)。应用 BASIC 程序 SPASSOC.BAS 确定方差比率检验、卡方和关联指数(表 11.7)。表明存在总体负关联($VR=0.44$)。但 $W=4.45$, 落入 3.96—18.3 的范围内, 所以我们接受八树种间无关联的零假设。如果我们查看成对的比较就会发现, 计算出的全部卡方值都是有偏的。似乎很有可能存在着实际上的负关联(种对 1—8, 2—6 和 2—8)和正关联(种对 6—8 和 7—8), 如八树种的网状图所示(图 11.5)。这些结果表明需要进一步的研究来证实或否定这些可能的关联。

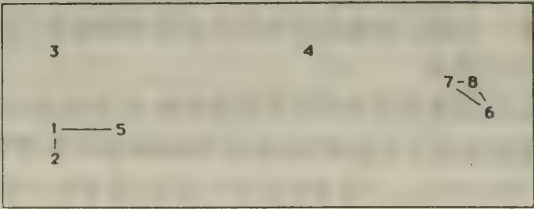


图 11.5 网状图示在威斯康星南部山地森林中八个树种间的关联格局。由直线连结的种具正的关联

表 11.6 在威斯康星南部 10 个山地森林取样单位中八个树种的(a)多度和(b)出现—不出现的生态数据矩阵(见 Peet 和 Loucks 1977)

(a)多度											
		取样单位									
种名	No.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Bur oak	(1)	9	8	3	5	6	0	5	0	0	0
Black oak	(2)	8	9	8	7	0	0	0	0	0	0
White oak	(3)	5	4	9	9	7	7	4	6	0	2
Red oak	(4)	3	4	0	6	9	8	7	6	4	3
American elm	(5)	2	2	4	5	6	0	5	0	2	5
Basswood	(6)	0	0	0	0	2	7	6	6	7	6
Ironwood	(7)	0	0	0	0	0	0	7	4	6	5
Sugar maple	(8)	0	0	0	0	0	5	4	8	8	9

(b)出现—不出现											
		取样单位									
种名	No.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Bur oak	(1)	1	1	1	1	1	0	1	0	0	0
Black oak	(2)	1	1	1	1	0	0	0	0	0	0
White oak	(3)	1	1	1	1	1	1	1	1	0	1
Red oak	(4)	1	1	0	1	1	1	1	1	1	1
American elm	(5)	1	1	1	1	1	0	1	0	1	1
Basswood	(6)	0	0	0	0	1	1	1	1	1	1
Ironwood	(7)	0	0	0	0	0	0	1	1	1	1
Sugar maple	(8)	0	0	0	0	1	1	1	1	1	1

11.6 种间关联的附加讨论

Pielou(1977)区别出完全的(complete)和绝对的关联(absolute association)。完全关联表现的情形是一个种总是随第二个种的出现而出现,但第二个种可以在没有第一个种出现的一些SU_s中出现(情况 1,表 11.8)。这种情况出现在当第二个种较常见时(即格 b 或格 c 等于零,但不同时等于零)。绝对联系存在于当无论那个种都不会在另一个种未出现时出现(即 b=c=0;情况 2,表 11.8)。需要一个能区别绝对关联与完全关联的指数,这个指数可以象表 11.8 中 Ochiai,Dice 和 Jaccard 指出那样。

Ochiai,Dice 和 Jaccard 指数是非概率关联指数的例子。仍提出许多概率指数(列联表,卡方几率)以测度 a-E(a)的数量离开 0 的程度(参见 Hubalek 1982),正象我们在本章前面讨论过的,关联有两个明显不同的内容:(1)显著性的统计检验(使用卡方检验统计量),(2)测度关联程度的指数。通常涉及关联的两个内容的概率指数的例子是 Yule 的(1912)指数(YI):

$$YI = \sqrt{\frac{X^2}{N}} \quad (11.17)$$

表 11.7 在威斯康星南部十个山地森林取样单位中八树种的种间关联指数和检验统计量

种对		多个种						
		VR, 总体关联指数=0.45						
		检验统计量 W=4.45						
		种对						
种对		关联类型 ^a	卡方		Yates 卡方	相关指数		
			值	偏差		Ochiai	Dice	Jaccard
1 2		+	4.44	*	2.10	0.82	0.80	0.67
1 3		+	1.67	*	0.05	0.82	0.80	0.67
1 4		-	0.74	*	0.05	0.68	0.67	0.50
1 5		+	3.75	*	1.28	0.87	0.86	0.75
1 6		-	4.44	*	2.10	0.33	0.33	0.20
1 7		-	3.40	*	1.41	0.20	0.20	0.11
1 8		-	6.67	*	3.75	0.18	0.18	0.10
2 3		+	0.74	*	0.05	0.67	0.61	0.44
2 4		-	1.67	*	0.04	0.50	0.46	0.30
2 5		+	1.67	*	0.23	0.71	0.67	0.50
2 5		+	1.67	*	0.23	0.71	0.67	0.50
2 6		-	10.00	*	6.27	0.00	0.00	0.00
2 7		-	4.44	*	2.10	0.00	0.00	0.00
2 8		-	6.67	*	3.75	0.00	0.00	0.00
3 4		-	0.12	*	1.98	0.89	0.89	0.80
3 5		-	0.28	*	0.63	0.82	0.82	0.70
3 6		-	0.74	*	0.04	0.68	0.67	0.05
3 7		-	1.67	*	0.04	0.50	0.46	0.30
3 8		-	1.11	*	0.00	0.60	0.57	0.40
4 5		-	0.28	*	0.63	0.82	0.82	0.70
4 6		+	1.67	*	0.05	0.82	0.80	0.67
4 7		+	0.74	*	0.05	0.67	0.62	0.44
4 8		+	1.11	*	0.00	0.74	0.71	0.56
5 6		-	1.67	*	0.23	0.58	0.57	0.40
5 7		-	0.10	*	0.23	0.53	0.50	0.33
5 8		-	2.50	*	0.63	0.47	0.46	0.30
6 7		+	4.44	*	2.10	0.82	0.80	0.67
6 8		+	6.67	*	3.75	0.91	0.91	0.83
7 8		+	6.67	*	3.75	0.89	0.89	0.80

a: 符号表示种间关联的方向

*: 未校正的卡方值是有偏差的, 或者因为(1)2×2表的任何格子中的期望频度(1和或(2)不止两个格子中的期望频度(5)。

其中 χ^2 是在方程 (11.7) 所计算的卡方值。Yule 指数的范围是从 -1 (当 $a=b=0$) 到 +1 (当 $b=c=0$), 并常常被称为离散二变量 (出现一不出现数据) 的点协相关系数 (Pielou

1977)。从表 11.8 可以看到：YI 区分完全关联和绝对关联。但是，与 Ochiai, Dice 和 Jaccard 指数不同，这个指数在计算中包含 d 格的值，而在情况 3 (表 11.8) 两个种不一块出现时，YI 的值等于-0.50。

表 11.8 完全关联 (情况 1)、绝对关联 (情况 2) 和无关联 (情况 3) 的四种关联测量值

情况 (1)

		种 B		
		出现	不出现	
种 A	出现	25	25	50
	不出现	0	25	25
		25	50	75

情况 (2)

		种 B		
		出现	不出现	
种 A	出现	25	0	25
	不出现	0	25	25
		25	25	50

情况 (3)

		种 B		
		出现	不出现	
种 A	出现	0	25	25
	不出现	25	25	50
		25	50	75

	情况		
指数	(1)	(2)	(3)
Ochai	0.71	1.0	0.0
Dice	0.67	1.0	0.0
Jaccard	0.50	1.0	0.0
Yule	0.50	1.0	-0.5

YI 作为关联强度的测度还有许多理论上的和实际上的限制 (见 Janson 和 Vegelius 1981, 和 Hubalek 1982)。但在实际上, Hubalek 发现 YI 与其他指数相比表现相当令人满意, 并对 OI、DI 和 JI 的使用提供了有益的补充。

在前面我们讨论了根据出现-不出现数据查明物种关联的 Schluter (1984) 方差比率检验。Schluter 还把他的方法扩展到使用物种的密度数据。McCulloch (1985) 给出了 Schluter 关联检验与某些标准统计检验之间的关联。

11.7 总结和推荐

1. 种对间关联的检验是：是否它们在一系列的取样单位上的共同存在是独立的。如果它们一块出现的次数比期望的更频繁，它们就具有正的关联；如果它们的共同出现少于在独立情况下所期望的，则它们具负的关联。

2. 有两个不同的内容用于确定种间的关联。第一，有一个种独立性的统计检验，其二，有一个关联强度程度的测度。我们建议：种间关联的研究应包括这两项内容(11.2.1和11.2.2节)。

3. 虽然可以用自然的和人为的 SU_s 来获得种出现—不出现数据，但本章提供的方法最好使用自然 SU_s 。人为 SU_s 在估计 SU 的大小对物种频度的影响上存在许多问题。

4. 在生态学文献中有许多让人迷惑的关联指数。我们介绍 Ochiai, Dice 和 Jaccard 指数是因为这些指数在应用上简单和易于理解。进而，根据关联指数特性的广泛研究，已证明这三个指数具优良性能(11.2.2节)。

5. 当目的是同时在一大群物种中评估关联的存在和不存在时，我们推荐 Schluter 的方差比率检验(11.2.3节)。

6. 确定关联的存在或不存在不能使我们了解可能的原因。我们建议使用大的样本容量；学生们在解释结果时总要格外留心才行。

第十二章 种间协变

当一个样本含有物种多度（如生物量）的定量测度时，即可估算出种间多度的协变（Covariation）。如果两个种的多度趋向于一致增加和减少，则这两个种是以相同的方式对相同的环境因子作出反应吗？如果当其他种减少时一个种的多度格局总是增加，则存在某些可能的负相互作用吗？在本章中介绍使用相关系数来测度成对种的多种数据协变的相对强度。

12.1 基本方法

在第十一章中，我们通过检验共存种对的独立性来考查种间关联。这完全根据出现一不出现数据。如果它们的共同出现不是独立的，则确定物种间具正的或负的关联。如种的多度数据可用（例如密度、百分盖度、生物量），涉及物种亲和性的接踵而来的问题即被提出。如果一个种的多度高，这对另一个种的多度能或多或少起到减少或增加的作用吗？这类协变问题与第 11 章提出的那些问题完全不同。事实上，两个种因在取样单位（SU_s）中同时出现而表现出强的正关联，但可能具强的负协变（即当一个种的多度增加，而其它种减少）。由于这个原因（Hurlbert (1969) 先前曾提出的一个论点，认为重要的是在物种关联与物种协变间确定一个明显的区别。不幸的是这两个术语常被不适当地互换了。术语相关（Correlation）是常用的，但在许多生态学场合中这个术语被保留下来作为隐含（并测度）两个种如何协变变化的意义。

在本章中，将用相关系数在一系列 SU_s 中检查种对的多度格局。SU_s 可以是自然单位，也可以是人为的，虽然在后一情况下一定要谨慎。在 11.2 节中关于使用人为 SU_s 的讨论在此同样有用。

值得强调的还有：两个种的多度格局间在统计上显著相关的确定并不能说明之所以如此的可能的潜在原因。但是，确定显著的种间协变对导出合适的现象来解释这样的格局很有帮助，从而可以导致进一步的实验研究。

12.2 步骤

设第 i 物种在 N 个 SU 中的多度由向量 $y_i = [Y_{i1}, Y_{i2}, \dots, Y_{iN}]$ 来表示，第 k 物种则由向量 $y_k = [Y_{k1}, Y_{k2}, \dots, Y_{kN}]$ 表示。这两个种间的正相关意味着当一个种的多度在一个 SU 中增加时则在此 SU 中另外的种也有相应的多度增加。同样，负相关则意味着由于一个种的增加而使另外的种减少。如果向量 y_i 和 y_k 是各自取自正态分布的样本（所谓的双变的正态分布 Bivariate normal distribution），则我们可以计算 Pearson 积矩相关系数（Pearson product-moment correlation coefficient）并检验事实上存在显著相关假设。如果双变分布不是正态的，我们可以计算非参量系数——Spearman 秩相关（Spearman rank correlation）。两种系数的使用足以断定物种多度向量间存在线性相关关系。

12.2.1 Pearson 相关

步骤 1. 提出零假设。 零假设为物种多度非相关。

步骤 2. 计算 Pearson 系数。 按下式计算第 i 与第 k 种间的 Pearson 相关 (r):

$$r(i, k) = \frac{\sum y_i y_k}{\sqrt{\sum y_i^2 \sum y_k^2}} \quad (12.1)$$

其中 $\sum y_i y_k = \sum_{j=1}^N Y_{ij} Y_{kj} - [(\sum_{j=1}^N Y_{ij})(\sum_{j=1}^N Y_{kj})/N]$

$$\sum y_i^2 = \sum_{j=1}^N Y_{ij}^2 - [(\sum_{j=1}^N Y_{ij})^2/N]$$

$$\sum y_k^2 = \sum_{j=1}^N Y_{kj}^2 - [(\sum_{j=1}^N Y_{kj})^2/N]$$

给定 Y_{ij} = 在第 j 个 SU 中第 i 种的多度

Y_{kj} = 在第 j 个 SU 中第 k 种的多度

方程 (12.1) 的分子是种 i 和种 k 间的协方差, 即多度如何共同协变的测度。如果这个协方差等于 0, 则 $r=0$, 即两个种间的相关等于 0。换言之, 一个种多度大小的变化并不意味着另外的种多度也相应变化。种的多度可以以正或负的形式协变, 且协变越剧, 相关越强。Pearson 相关系数的取值范围是从 -1 (完全负相关) 到 +1 (完全正相关), 且系数无单位。

步骤 3. 检验零假设。 按方程 (12.1) 用样本数据算出的 r 值是种群相关系数 ρ 的估值。为了检验两个向量 y_i 和 y_k 间显著相关的存在, 我们检验 $\rho=0$ 的零假设, 相对的选择是 $\rho \neq 0$ 。换言之, 例如, 如果 $r=0.47$ (如由随机样本确定的), 总体中 y_i 和 y_k 间实际存在相关或 r 值简单地是偏离 0 的机率吗?

为了检验零假设, 用 Rohlf 和 Sokal (1981, 表 25, 第 168 页) 给出的, 具 $df=N-2$ 和 5% 的概率水平的临界值与 r 的绝对值比较。如果 $|r|$ 超出这个临界值, 则拒绝零假设。

作为一种亲和性指数 Pearson 的 r 具有一些性质限制了它本身的有用性 (Boesch 1977): (1) 它趋向于夸大数据中很大的数值对总体的重要性, (2) 当数据中有许多 0 时它可能产生假相关 (见 12.5 节), (3) r 的显著性检验呈潜在的正态频率分布, (4) r 在 SUs 中种多度间呈线性相关。

12.2.2 Spearman 的秩相关

通常在 y_i 和 y_k 中双变种群数据远非正态分布。如果这样, 前一节中的步骤则不适宜了, 故我们求助于协变的非参量测度。一个常为生态学家使用的流行系数是 Spearman 秩相关:

步骤 1. 提出零假设。 零假设为“秩化的”种多度非相关。

步骤 2. 计算 Spearman 系数。 顾名思义, 第一项任务是按从最大值到最小值的顺序在 y_i 和 y_k 中排列多度数据。用 $N=5$ 和种 i 和 k 的简单例子来解释则易于理解。如果多度向量是

$$y_i = [8, 6, 3, 0, 2]$$

$$y_k = [2, 14, 0, 6, 6]$$

而秩化的向量是：

$$y_i(\text{秩化的}) = [5, 4, 3, 1, 2]$$
$$y_k(\text{秩化的}) = [2, 5, 1, 3.5, 3.5]$$

注意： y_i 的秩化是简单的，而在 y_k 中则有一点麻烦，即“相对峙”，因为有两个多度值为 6。当出现相对峙时，每个相对峙数被给定一个秩，这个秩是在无相对峙情况出现的条件下给定的秩的平均值。故在上例中，各个 6 都被指定为 3.5 秩，这是 3 和 4 的平均值，即如果未出现相对峙被指定的秩。

然后将 y_i （秩化的）和 y_k （秩化的）代入方程（12.1）来计算以 r_s 表示的 Spearman 秩相关系数。与 Pearson 的 r 一样， r_s 无单位且取值范围从 -1 到 +1。

步骤 3. 检验零假设。对于秩化的种多度无相关的零假设，我们使用与 Pearson r 同样的检验和基本原理。只要 $N > 10$ ，就可以采用具相对小误差的同样临界值表（Sokal 和 Rohlf 1981）。如果 $N \leq 10$ 则应参考表 12.1。

表 12.1 在小样本中 Spearman 秩相关的显著性水平。如果 $|r_s|$ 的计算值大于样本容量为 N 的表上数值则相关在给定的概率水平上是显著的（引自 Snedecor 和 Cochran 1973）

样本容量(N)	概率水平	
	P=0.05	P=0.01
≤4	—	—
5	1.00	—
6	0.886	1.00
7	0.750	0.893
8	0.714	0.857
9	0.683	0.833
10	0.648	0.794
≥11	a	a

a 使用表 A 11, 557 页, Snedecor 和 Cochran (1973) 或 Rohlf 和 Sokal (1981) 表 25, 168 页

12.2.3 多物种间的协变

如同在前一章的种间关联那样，我们的兴趣常常在于以成对组合的许多个种是如何协变的。许多种之间的协变格局有两种作图方法：

作图 1. 协变比较矩阵。与种间关联一样，种间协变的结果可以矩阵的形式总结于图 12.1 所示的协变比较矩阵。

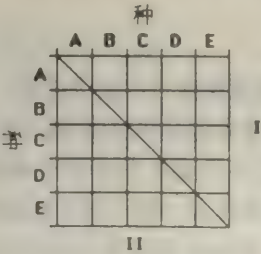


图 12.1 成对物种协变矩阵。右上三角阵(I)可用于表示相关而左下的三角阵(II)可用于表示协方差

作图 2. 协变网状图。种间协变的格局可以网状图的形式人为地表示(见 11. 2. 3 节)。种与种之间的距离在网状图中反映它们正协变的相对程度,即那些具正的和显著协变的种离得近,而那些负协变的种离得较远。例如,给出五个种 A、B、C、D 和 E 之间成对的种间协变,图 12. 2 表明种 B 和 D 可与种 A、C 和 E 分开成组。

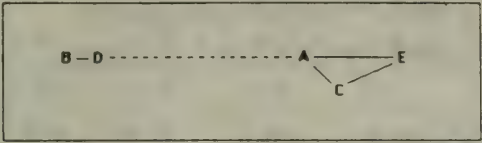


图 12.2 物种协变格局的网状图。种 B 和 D 同种 A、C 和 E 一样具正协变(直线),但这些组由负协变划分(虚线)

12.3 举例:计算

两个种 i =羊茅属(*Festuca*)和 k =薊属(*Cirsium*)的百分盖度值获自英格兰白垩草原上的九个相邻样方(SUs) Greig-Smith 1983):

$$y_i = [88.4, 84.8, 74.1, 73.2, 60.7, 57.1, 55.4, 43.8, 41.1]$$

$$y_k = [8.9, 3.6, 19.6, 11.6, 31.3, 23.2, 18.8, 32.1, 33.0]$$

仅根据出现—不出现情况,我们注意到种间完全关联(即总是共同出现;见第 11 章)。利用这些多度数据,我们可以计算在盖度上是否存在显著的协变。根据两个多度向量的视觉印象,有薊属植物的盖度随羊茅属植物盖度的减少而增加的趋势。

12.3.1 Pearson 相关

Pearson 相关系数是[方程(12.1)]

$$r(i,k) = \frac{\sum y_i y_k}{\sqrt{\sum y_i^2 \sum y_k^2}} = \frac{-1,285.06}{\sqrt{(2,278.81)(917.18)}} = -0.89$$

其中 $\sum y_i y_k = [(88.4)(8.9) + \dots + (41.1)(33.0)]$
 $- \{[(88.4 + \dots + 41.1)(8.9 + \dots + 33.0)]/9\}$
 $= 10,421.95 + \dots + [(578.6)(182.1)/9] = -1,285.06$
 $\sum y_i^2 = (88.4^2 + \dots + 41.1^2) - [(88.4 + \dots + 41.1)^2/9]$
 $= 39,476.36 - [(578.6)^2/9] = 2,278.81$
 $\sum y_k^2 = (8.9^2 + \dots + 33.0^2) - [(8.9 + \dots + 33.0)^2/9]$
 $= 4,601.67 - [(182.1)^2/9] = 917.18$

为了检验 Pearson 相关的显著性,我们把 r 的计算值(0.89,符号忽略)与获自 Rohlf 和

Sokal(1981,表 25,168 页)或 Snedecor 和 Cochran(1973,表 A 11)的临界值 0.67(df=7,P=0.05)比较。由于 $|r|$ 超出临界值,我们拒绝种多度非相关的零假设,并得出结论说在这两个种的盖度值中存在强的负相关。

12.3.2 Spearman 相关

在本例中样本容量很小(<10),故正态分布的假定即为存疑。通过先按从最高到最低盖度值的顺序将向量 y_i 和 y_k 秩化的方法计算非参量的 Spearman 秩相关:

$$y_i(\text{秩化}) = [9, 8, 7, 6, 5, 4, 3, 2, 1]$$

$$y_k(\text{秩化}) = [2, 1, 5, 3, 7, 6, 4, 8, 9]$$

现在根据方程(12.1)

$$r_{s(i,k)} = \frac{\sum y_i(\text{秩化})y_k(\text{秩化})}{\sqrt{\sum y_i^2(\text{秩化})\sum y_k^2(\text{秩化})}} = \frac{-50}{\sqrt{(60)(60)}} = -0.83$$

$$\begin{aligned}\text{其中 } \sum y_i(\text{秩化})y_k(\text{秩化}) &= [(9)(2) + \dots + (1)(9)] \\ &\quad - \{[(9 + \dots + 1)(2 + \dots + 9)]/9\} \\ &= 175 - [(45)(45)/9] = -50\end{aligned}$$

$$\begin{aligned}\sum y_i^2(\text{秩化}) &= (9^2 + \dots + 1^2) - [(9 + \dots + 1)^2/9] \\ &= 285 - [(45)^2/9] = 60\end{aligned}$$

$$\begin{aligned}\sum y_k^2(\text{秩化}) &= (2^2 + \dots + 9^2) - [(2 + \dots + 9)^2/9] \\ &= 285 - [(45)^2/9] = 60\end{aligned}$$

由于 $N < 10$,我们将计算值 $r_s = -0.83$ 与临界值 0.683(表 12.1; 7df; $P=0.05$)比较,并得出结论,认为 r_s 是显著的(在 $P=0.05$,不是在 $P=0.01$ 水平上)。

注意,当两个向量按指定的秩排列均相等时(如本例),分母项相等,仅需计算一项,即

$$\sum y_i^2(\text{秩化}) = \sum y_k^2(\text{秩化})$$

12.4 例子:威斯康星森林

在表 11.6a 中给出威斯康星南部十块高地林地上八种树木的相对密度组。为了对每一成对组合的种的协变值问题加以解决,使用 BASIC 程序 SPCOVAR. BAS 来计算 Pearson 和 Spearman 相关系数。结果总结于表 12.2。

物种协变网状图显示出物种协变的明显格局(图 12.3)。具较低数量(演替中的种)的树种有与具高数量(森林顶极树种)的树种呈负协变的趋势。

在 28 个成对的组合中使用 Pearson 相关有 10 对相关是显著的,但用 Spearman 秩相关则仅 6 对。但在此要谨慎地注意,样本容量小,故这些密度数据(表 11.6a)极不可能属于正态分布。给定这个条件后非参量的 Spearman 相关是适用的协变测度。还有在此数据组中有许多 0。在 12.5 中我们讨论了在此情况下可做的抉择:(1)删去稀有种或不常出现的种和/或(2)去掉全部双零配对。关于后者,考虑种 7(铁木)和种 8(糖槭)的情况。在计算 0.74 的 r_s 值中包含了

五个成对的零配对(表 12.2)。如果我们在计算中删除这些零,则种 7 和 8 的 r 值即为 -0.21 。所以,学生们应注意在具小样本容量的数据组中具零的潜在问题。

表 12.2 由程序 SPCOVAR. BAS 计算出的树种(SPP)多度数据(表 11.6a)的 Pearson (右上角)和 Spearman(左下角)相关系数

SPP	Pearson 积矩相关							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	1.00	0.67*	0.22	-0.22	0.30	-0.79**	-0.48	-0.84**
(2)	0.66	1.00	0.38	-0.57	0.02	-0.90**	-0.64*	-0.74*
(3)	0.16	0.26	1.00	0.13	0.12	-0.53	-0.71*	-0.62
(4)	-0.02	-0.56	0.20	1.00	0.01	0.35	0.02	0.05
(5)	0.32	-0.09	0.10	0.11	1.00	-0.30	0.06	-0.32
(6)	-0.77*	-0.86**	-0.45	0.40	-0.30	1.00	0.74*	0.90**
(7)	-0.48	-0.62	-0.71	0.01	0.07	0.63	1.00	0.77**
(8)	-0.85*	-0.73*	-0.58	0.01	-0.28	0.83*	0.74*	1.00

Spearman 秩相关

* =在 5%概率水平上显著
* * =在 1%概率水平上显著

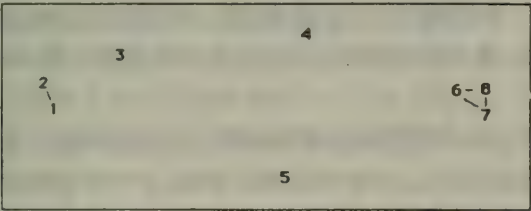


图 12.3 威斯康星南部山地森林中八个树种的网状图。由直线连结的种间具正的协变

最后,由于种的组合有 $S(S-1)/2$ 对,对每一种对使用一维显著性检验可能导致严重的解释错误(Legendre 和 Legendre 1983)。物种不是独立的,故检验也不是独立的。这个问题与我们在 11.2.3 节当使用卡方法检验种间关联时提出的问题一样。如在下一节中讨论的那样,多维检验是可采用的。

12.5 种间协变的附加讨论

生态数据组常常以存在许多零为特点。回忆第 11 章选择关联指数的条件之一就是不计双零的配对(11.2.2 节,条件 5);否则,可能导致严重曲解。在相关系数的计算中,零被处理为“定量的”数据点。为了避免曲解,我们提出三点(Legendre 和 Legendre 1983):(1)删除不常见的种,这对寻找有意义的亲和性的影响必然最小;(2)删除全部零,即把它们当做丢失数据处理;

或(3)只删除双零配对。根据与目的相关的数据特性,我们推荐第1和第3两点。但是如果目的是关于物种由竞争影响导致的格局间相关关系的,则仅含有两物种的 SUs 可用(Hurlbert 1969)。

除利用种对计算种间协变外,我们可以提出这样的问题,即许多 S 种的多度在群落中的 N SUs 间是否显著协变。第11章中谈及的方差比率检验可用来回答这一问题。Schluter(1984)把使用多度数据情况下出现—不出现数据的方差比率检验加以归纳。所得方差比率可用作物种协变的指数:一个显著大于1的值指示物种间趋于正协变,相反,显著小于1的值指示物种间具负协变。Schluter 提供了一个接近卡方分布的检验统计量(W)。

用于测度物种协变的其他系数包括 Kendall's (Greig-Smith 1983)和 Morisita (1959)相关系数。后者有利于比较广泛使用的 Pearson 积矩相关(Hurlbert 1969)。Morisita 相关系数相对独立于 SUs 中平均物种多度和多样性(Wolda 1981)。

在前一节中,我们注意到:在物种的成对组合间存在着许多相关关系的显著性个别检验的问题。由于物种不是独立的,显著性检验也不是独立的,故应使用多维的,同时的检验(Legendre 和 Legendre 1983)。Bartlett(1954)给出了一个统计量用于检验 $S(S-1)/2$ 相关系数等于零的零假设,Johsson 和 Wichern (1982)描述了一个多维过程,旨在检验是否相关系数的总数显著地不同于其他。

12.6 总结和推荐

1. 如果物种的多度数据可用(如盖度、生物量、数量),我们就可以确定种间这些数据的协变。这种类型的分析与在第11章中提到的那些种间关联技术截然不同,种间关联只以物种的出现—不出现数据为根据。

2. 相关系数可用于测度两种间多度数据协变的程度。正的相关蕴含着当一个种的多度增加时,另外的种有一相应的增加的意义。对于负的相关来说,一个种的增加就意味着另一种的减少。当然,确定相关关系的存在并不意味着存在因果关系。

3. 如果数据取自双变正态分布,则可使用 Pearson 相关系数(12.2.1节)。如果数据是非正态的,则引用非参量的 Spearman 秩系数(12.2.2节)。两个系数均假定被比较的两物种的多度间存在线性相关关系。

4. 许多群落生态学数据组以具许多零为特点,即许多 SUs 缺少某些种(多度=0)。我们建议学生们在计算相关系数时,首先删除不常见的种,然后删除双零的配对(两个种的多度均为零的 SUs)。不这样做可能导致无意义的(假的)相关(12.2.3节和 12.4节的例子)。

5. 由于物种不是独立的,故当解释大量的物种成对组合间相关关系的统计显著性时要格外谨慎;故统计检验也是如此(12.5节)。

第五部分 群落分类

第十三章 背景

给出一系列实体和计算他们相互之间相似性的某种方法,我可以把分类定义为基于相似性基础上的,这些实体的分组或聚类。在寻找 Sokal(1974)所谓的“自然”系统时,分类在许多科学领域内起着十分重要的作用。一个自然系统可以被认为是那些各种各样的导致实体进行排列的过程的反映。例如,在生态学中这个“自然”系统可能就是进化过程的最终结果。

生态学群落分类的第一步就是取样。采用人为取样单位(SUs)或自然取样单位,可以得到定性的和定量的各种数据。这些数据可以包括种类出现的列表或者种丰富度的某种指标(密度、频度、盖度、生物量)。第二步是用某种方法计算所有取样单位对之间的生态学相似性,以便确定他们之间的相似性或相异性的大小(我们把它称为 Q-方式,见第九章)。最后,依据它们的相似性把取样单位(实体)进行分组。每一组内的取样单位之间应该有若干个共同特征把它们和其它组内的取样单位分离,其目的是表明各取样单位之间的相互关系,并希望简化这些关系,以便可以对已经存在实体的分类做出一般性的说明。

Able 和 Noon(1976)的研究是一个潜在分类问题的很好示例。他们的目的是在纽约的阿第伦达克山上,沿高度梯度描述鸟类群落的结构。他们沿着一个海拔从 400m 到 1400m 的带状调查剖面上发现了 44 种鸟类。有些种类的鸟出现在沿梯度分布的所有取样单位内;其它的种类则仅见于一个狭窄的高度范围内。沿着高度梯度,在几个地段(交错群落)上,植被结构发生了较大的变化,因此形成了一些鸟类分布的自然上限和下限。在他们的研究中基于种类丰富度数据,取样单位的分类将反映出全部 44 种鸟类的连续分布和间断分布。这儿我们感兴趣的自然系统,在广义上说就是沿高度梯度反映大多数鸟类丰富度的取样单位的分组。

当然,同质群体没有必要进行分类。而且,在第十五、十六章内介绍的一些技术甚至可以把随机数据集分类,因此在解释这些结果时我们必须小心。只因为它可能把一个数据集进行分类,一个不合逻辑的分类将不会产生一个有生态学意义的解释。Goodall(1970),Ratcliff 和 Pieper(1982),Sokal (1974) 和 Whittaker 1978a, b)查阅回顾了各种分类理论的基本原理;我们极力推荐,对分类非常感兴趣的同学读一读这些论文。

在生态学发展的早期,群落的分类很大程度上是凭直觉,基于主观决定和定性描述。最近则倾向于基于定量数据的客观分类方法。在以后的几章中我们将检验这些客观方法中的一些方法。所有分类研究中存在一定的主观性是不可避免的,鉴于一个给定的分类方法只能产生唯一的结果,而一个有选择性的方法可以产生不同的结果,因此,研究者必须做出主观抉择。

此处对以后各章将用到的术语简要解释如下:

1. 分类可以有等级的，也可以是网状的。正如它的名字所表明，在一个等级分类中任何分类较低等级上的组都是较高等级中组的次级组。在网状分类(我们不做讨论)中，各组之间是独立的，而不是按等级排列，它们之间以网状系统连接。

2. 分类可以是分裂的，也可以是聚合的。在分裂分类中，所有取样单位根据他们的相似性进行划分，再划分，直到最后一个分组完成(画出一个倒的树状图)。在聚合分类中，顾名思义，各个取样单位之间进行结合，再结合形成更大的取样单位组(树状图)(直到都融合成一个组)。

3. 分类可以是单元的，也可以是多元的。在一个单元分类中，任意两个取样单位或任意两组之间的相似性只是根据一个变量值求得，例如，一个种的有无。在多元分类中，任意两个取样单位或任意两组之间的相似性是根据多个变量的总的相似性求得，例如，多个种的丰富度。

在分类方法的下一步处理中，我们即采用定性数据(例如，种的有无或二态特征数据)也采用定量数据(例如，丰富度或有序多态特征数据)。有关 Q-方式分析的生态学相似指数将在第十四章中介绍。在第十五章中将介绍一般的关联分析技术;这是一个单元分裂分类模式，它基于取样单位内种的有无。在第十六章内我们将介绍多元聚类分类技术，一般称为聚类分析，该方法基于取样单位内种类丰富度的定量数据。

13.1 矩阵展示

在研究种间关系时(第十、十一、十二章)，生态学数据矩阵是按行考虑的(图 9. 2)，这是一个 R-方式分析。在分类研究中，生态学数据矩阵即可按行考虑，也可按列考虑，这是 Q-方式分析(图 13. 1)。这两种情况的目的是相同的，即把取样单位进行分类。

观察 R-方式分析(种)和 Q-方式(取样单位)分析之间的关系另一种方法是以几何方法概括这些关系;当涉及到 R-方式和 Q-方式的研究时，导致了多维空间(Hyperspace)一词的产生(Williams 和 Dale 1965)。种的多维空间被定义为 S 维的，也就是说取样单位内的 S 个种中的每一种代表一维(显然，我们无法画出超出三维的 S 维空间)。

		取样单位										
		1	2	3	4	5	N	
种	a	■	■	■	■	■						
	b	■	■	■	■	■						
	c	■	■	■	■	■						
	.	■	■	■	■	■						
	s	■	■	■	■	■						
因子	w											
	x											
	.											
	.											
	z											

图 13.1 阴影部分表明计算 Q-方式相似的生态学数据矩阵的形式。我们感兴趣的是成对的 SU 之间的相似性

然后根据取样单位内每个种的相对丰富度把各取样单位定位于这个 S 维种(Species)空间内。该 S 维(species)空间内各取样单位之间的距离代表它们之间相互的相似性(或它们之间的相

异性，第十四章)。图 13.2 给出的示例表明，两个种的二维空间内三个取样单位(Q-方式)的位置。

另一方面，取样单位多维空间被定义为 N 维的，样本中 N 个取样单位中的每一取样单位代表一维。依据它们的丰富度，各个种被定位于这个 N 维空间内；在这个空间内，两个种相距越近，它们在取样单位内的丰富度就越相似。图 13.2 画的是两个种在两个取样单位空间内的位置(R-方式)。这种类型的空间表示更具有人为性，因为取样单位轴上的值是取样单位内种的丰富度((Legendre 和 Legendre 1983)。

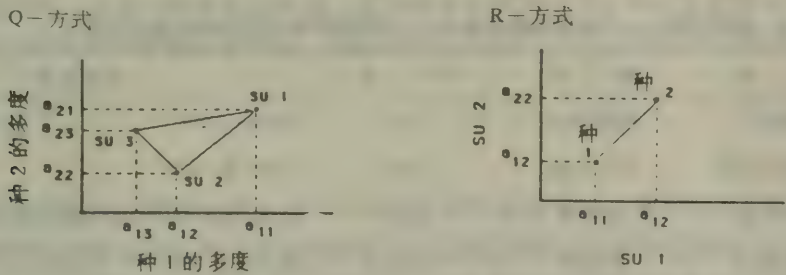


图 13.2 从几何意义上考虑 Q-方式和 R-方式分析。Q-方式是种空间内取样单位的代表，R-方式是取样单位空间内种的代表。注意 a_{ij} 是第 j 个取样单位内第 i 种的丰富度。
摘自 Legendre 和 Legendre (1983)。

最后，我们回忆一下一个取样单位和一个样本之间的区别。一个样本由一系列取样单位组成(表 1.1 给出了一个示例)。由于生态学群落的巨大多样性，学生应该知道生态学数据矩阵内的列即可表示单独的取样单位，也可表示样本。一些文献中的示例对此做了最好的说明。

在研究澳大利亚昆士兰莫顿湾水底生物群落时，Stephenson 等人(1970)对水底进行了大规模采挖取样。他们的取样单位是一个挖掘器，其口径为 $84 \times 29\text{cm}$ ，网孔 7.5cm ，切口刃斜角为 25° ，全长 2.5m 。这些特殊的规格十分重要，因为这些规格中任意一个发生变化都很可能会影响底栖动物的采集。他们的船速保持在 0.6km/小时 ；采集进行两分钟。把采集到的样品进行分类，对采集到的大型海底动植物进行测定。他们在整个海湾的 400 个采集点上收集到的 355 种生物，他们的生态学有一无数据矩阵是 355 行(种)乘 400 列(取样单位)的，对这些数据即作了 R-方式分析，也做了 Q-方式分析。

在下一个示例中，数据矩阵的列代表各个样本，而不是单独的取样单位。Huhta(1979)从 1962 年到 1965 年，又在 1968 年对芬兰赫尔辛基北部的未扰动并除去森林的地段的土壤节肢动物群落的组成进行了检测。在每一森林内每两个月进行一次取样。一次取样包括 4 个 $25 \times 25\text{cm}$ 的土壤和碎屑样方，把它们带回实验室，对节肢动物的总种数和它们的相对丰富度进行测定。这些每两个月一次取样的数据结合成年度样本进行分析。因此，Huhta 研究的数据矩阵其行是土壤节肢动物种，列代表年度样本(5 年)；矩阵中的数字是个体数量。Huhta 对这些数据进行了 Q-方式分析处理，来确定土壤节肢动物群落组成上的年度差异。

13.2 选择的文献

生态学中分类问题的研究非常之多。两个基本的分类技术将在以后几章中加以介绍；(1) 关联分析(AA, 第十五章)，(2) 聚类分析(CA, 第十六章)。聚类对策是有等级的，聚合的和多元的，它显然是被最普遍应用的方法。关联分析是有等级的，分裂的和单元的方法，不太常用。表 13.1 给出了分类研究示例的生态学文献。

表 13.1 应用关联(AA)分析或聚类分析(CA)研究群落分类示例的有关文献

地 点	群 落	方 法	参考文献
英国	白垩草地	AA	Gittins 1965
西北地区(加)	水生生物	AA	Vilks 等 1970
弗吉尼亚	落叶林	AA	Madgwick 和 Desrochers 1972
尼日利亚	萨王纳	AA	Kershaw 1973
澳大利亚	森林	AA	Ashton 1976
亚利桑那	荒漠草原	AA	Fish 1976
北海	海洋生物	CA	Stephenson 等 1972
澳大利亚	雨林	CA	Dale 和 Clifford 1976
北大西洋	海藻	CA	Lawson 1978
里科港	雨林	CA	Crow 和 Grigal 1979
纽约	落叶林	CA	Gauch 和 Stone 1979
英国	泥炭沼泽	CA	Clymo 1980
西北地区(加)	北极苔原	CA	Thompson 1980
澳大利亚	天然草地	CA	Foran 等 1986

第十四章 相似性函数

当研究群落结构的问题时,生态学家常常面临着对植物或(和)动物样本进行对比的任务。这些样本可以来自:(1)一个景观内的不同地点,例如,Able和Noon(1976)沿高度梯度对鸟类的分布进行的研究,或(2)同一地点的不同时间,例如Livingston(1976)对十二月和六月捕鱼数据的比较分析。在这一章中我们介绍一些使样本之间相似性、相异性定量化的几个相似性函数。样本的种类组成和数量越相似,它们的相似性越大,它们之间的生态学距离越小。

14.1 基本方法

正如Sneath和Sokal(1973)的定义,广义的相似函数是基于对一系列描述特征的观察,将两个实体之间的相似性或相异性定量化的函数。生态学家感兴趣的实体是取样单位(取样单位(SU)或样本,见第十三章),他们感兴趣的描述特征是种类丰富度的度量(例如,密度、生物量)。正如定义的那样,这些相似函数涉及到取样单位之间的Q-方式分析。

R-方式分析和Q-方式分析之间的区别已经在第十三章中作了介绍,用于这些不同分析方式的相似函数已在图9.1中作了说明。总地来看,Q-方式相似函数可以分成两类;(1)相似系数,(2)距离系数。相似系数的变化范围从最小的零(当一对取样单位完全不同时)到最大的1(两个取样单位完全相同时)。另一方面,相异系数则相反;当一对取样单位完全相同时,它们有一最小值0,当一对取样单位完全不同时,它们有一最大值(某些情况下是无穷大)。因此距离系数又叫相异系数。事实上,一个相似性指数总可以被表示成一个距离,虽然需要经过一个简单的变换,如:1-相似性(Legendre和Legendre 1983)。因此可以认为距离就是相似性(Sneath和Sokal 1973)。

不必说,相似函数的数量相当大。在这一章中我们的介绍仅限于Q-方式研究中较常用的几种相似性和距离的计算方法。但是这并不意味着前面提到的用于计算Q-方式相似性的一些统计概率指数不太好,相反,也许它们比我们将要介绍的几种方法更适用于某些数据集(见14.7节)。我们认为距离系数也许是最受群落生态学家欢迎的,它们在概念上以及在对群落数据的处理应用上都是最直接的。

14.2 步骤

14.2.1 相似性系数

相似性系数在生态学文献中显然是用得最多的指数(Legendre和legendre 1983)。这些指数只基于有(用1表示)无(用0表示)数据。例如,我们看一个有关3个取样单位中,3个种的有/无的例子:

种	取样单位		
	1	2	3
A	1	1	0
B	1	1	0
C	1	0	1

在 Q-方式分析中,我们对每一对取样单位(数据矩阵的列)的种类组成的相似程度感兴趣。相对于它们的总种类组分,两个取样单位的共有种越多,它们之间的生态学相似性就越大。在这个例子中,取样单位(2)中包括了取样单位(1)中三个种中的两种,但它和取样单位(3)没有共有种。

回忆一下,我们在 11.2.2 部分中介绍过三种基于有无数据,用于计算种群(R-方式分析,即按行考虑数据矩阵)之间关联程度的指数(Ochiai, Dice 和 Jaccard 指数)。同样这三个指数也可以用于计算取样单位之间的 Q-方式相似性。同学们应该注意,这些指数是我们唯一的即用于 Q-方式相似性(样本相似性)又用于 R-方式相似性(种间关系)计算的函数类型。上例中取样单位(1)和取样单位(2)之间的相似性分别用 Ochiai 指数[OI,方程 (11.10)], Dice 指数[DI,方程 (11.11)]和 Jaccard 指数[JI,方程 (11.12)]的计算结果是:

$$OI_{1,2} = \frac{2}{\sqrt{2} \sqrt{3}} = 0.82$$

$$DI_{1,2} = \frac{4}{4 + 0 + 1} = 0.80$$

$$JI_{1,2} = \frac{2}{2 + 0 + 2} = 0.67$$

因为我们已在第十一章中介绍过这些指数,因此在这一章中不再重述。

14.2.2 距离系数

距离的计算方法可分为如下三组:(1)E-组(欧氏距离系数),(2)BC-组(Bray-Curtis 相异性指数)和(3)RE-组(相对欧氏距离方法)。

以下的矩阵说明用于下面介绍的等式中: X_{ij} 代表第 J 个取样单位中第 i 种的丰富度。例如, $X_{4,3}$ 就是第三个取样单位中第四个种的丰富度。和前面介绍的一样,群落数据矩阵由 S 个种, N 个取样单位组成。

14.2.2.1 E-组距离

距离 1. 欧氏距离 Euclidean Distance(ED)。这种方法在计算欧氏空间内两个取样单位点 SU_j 和 SU_k 之间的距离的方程中是常见:

$$ED_{jk} = \sqrt{\sum_{i=1}^s (X_{ij} - X_{ik})^2} \quad (14.1)$$

因为 SU 内每个种的差值被平方后相加,因此 ED 加强了取样单位之间种丰富度的差值。

最后的距离值又由于总和开平方而缩小。ED 值的变化范围是 0 至无穷大,所有 E 组方法都如此。

距离 2. 平方欧氏距离 Squared Euclidean Distance (SED)。该方法只是 ED 的平方

$$SED_{jk} = \sum_{i=1}^s (X_{ij} - X_{ik})^2 \quad (14.2)$$

距离 3. 平均欧氏距离 Mean Euclidean Distance (MED)。MED 和 ED 相似,但最后的距离较小,因为其中使用了平均差异:

$$MED_{jk} = \sqrt{\frac{\sum_{i=1}^s (X_{ij} - X_{ik})^2}{s}} \quad (14.3)$$

距离 4. 绝对距离 Absolute Distance (AD)。这种方法是 S 个种绝对丰富度差值的总和:

$$AD_{jk} = \sum_{i=1}^s |X_{ij} - X_{ik}| \quad (14.4)$$

因为它是丰富度的差值之和,而没有平方,因此它不如前面三个距离那样强调取样单位之间的差值,因此得出的差值较小,而最后的距离则相对较大。这种距离方法在数值分类上被称为性状差值(Sneath 和 Sokal 1973)。

距离 5. 平均绝对距离 Mean Absolute Distance (MAD)。MAD 和 AD 相似,只是其中用了平均距离,而不是绝对距离:

$$MAD_{jk} = \frac{\sum_{i=1}^s |X_{ij} - X_{ik}|}{s} \quad (14.5)$$

MAD 和数量分类的平均性状差值相同(Sneath 和 Sokal 1973)。

14.2.2.2 BC—组距离。

这组只有一个由 Bray 和 Curtis(1957)首次引入生态学文献的指数。该指数在生态学家中始终很受欢迎。第一步是计算取样单位 J 和 K 之间的百分比相似性(Percent similarity—PS):

$$PS_{jk} = \left(\frac{2W}{A+B} \right) \times 100 \quad (14.6a)$$

此处:

$$W = \sum_{i=1}^s [\min(X_{ij}, X_{ik})]$$

$$A = \sum_{i=1}^s X_{ij} \text{ 和 } B = \sum_{i=1}^s X_{ik}$$

因此,第 J 个取样单位和第 K 个取样单位之间的 PS 是这样一个分数,其分子是每一对取样单位中每种的种丰富度 X_{ij} 和 X_{ik} 中最小者之和的二倍,其分母是两个取样单位中所有种的种丰富度之和。任何一对种丰富度完全相同的取样单位,则它们完全相似,也就是:PS=100%。

和 PS 互补的距离是百分比相异性(PD),按下式计算:

$$PD = 100 - PS \quad (14.6b)$$

PD 也可以在 0—1 范围内,计算方法如下:

$$PD = 1 - [2W/(A+B)] \quad (14.6c)$$

这点十分有用,因为这和由其它许多指数计算出的结果范围更为一致。在下面的计算中我们

将采用 PD 进行计算(14.6c)。

14.2.2.3 RE 一组距离。

该组中包含的距离指数表现为标准化或相对的尺度。

距离 7. 相对欧氏距离 Relative Euclidean Distance (RED)。该方法把每一取样单位内的每种丰富度进行了合并, 相对于总的取样单位丰富度的差值, 最后距离的大小是经过标准化的:

$$RED_{jk} = \sqrt{\sum_{i=1}^s \left[\left(\frac{X_{ij}}{\sum_i X_{ij}} \right) - \left(\frac{X_{ik}}{\sum_i X_{ik}} \right) \right]^2} \quad (14.7)$$

应用 Whittaker 的相对变换, 把 AD(方程 14.8)转变成 Orloci(1978)提出的 ED。RED 的变化幅度是 0 到 $\sqrt{2}$ 。

距离 8. 相对绝对距离 Relative Absolute Distance (RAD)。该方法是把 Whittaker 的相对丰富度校正用于 AD(同样, 相对欧氏距离“校正”了欧氏距离):

$$RAD_{jk} = \sum_{i=1}^s \left| \left(\frac{X_{ij}}{\sum_i X_{ij}} \right) - \left(\frac{X_{ik}}{\sum_i X_{ik}} \right) \right| \quad (14.8)$$

RAD 的变化范围是 0 到 2。

距离 9. 弦距离 Chord Distance (CRD)。该方法更重视取样单位内种的相对比例, 相应的则不太重视他们的绝对数量。从技术上看, 它的产生是通过利用方向余弦, 把取样单位投影到一个单位半径的圆上。这样投影之后, 它们之间的距离大小就是两个取样单位之间的 CRD。我们建议同学们读一下 Pielou (1984, P48)对此作的几何说明。弦距离由下式给出:

$$CRD_{jk} = \sqrt{2(1 - \text{ccos}_{jk})} \quad (14.9a)$$

此处弦余弦(Chord cosine—ccos)由下式计算得到:

$$\text{ccos}_{jk} = \frac{\sum_{i=1}^s (X_{ij}X_{ik})}{\sqrt{\sum_i X_{ij}^2 \sum_i X_{ik}^2}} \quad (14.9b)$$

注意, 当利用有一无数据时, CCOS 和 Ochiai 系数一样。象 RED 一样, CRD 的范围是 0 到 $\sqrt{2}$ 。

距离 10. 测地学距离 (Geodesic Distance (GDD))。该方法是各取样单位被投影到单位半径圆上后, 沿着圆弧的距离(而不是弦距离)。

$$GDD_{jk} = \arccos(\text{ccos}_{jk}) \quad (14.10)$$

GDD 的范围是 0 到 $\pi/2$ (即 0—1.57)。

根据前面介绍的相似性或距离的计算方法, 计算所有可能取样单位对之间的距离, 很容易得到一个距离(或相似性)值的 $SU \times SU$ 矩阵。检查这个矩阵可以很快发现任何两个我们感兴趣的取样单位之间的距离。群落分类的聚类对策就是在这个距离矩阵上进行的(第十六章)。我们在 14.6 部分中将给出一距离矩阵的例子。

14.3 举例:计算

为了说明取样单位之间距离大小的计算,我们将利用表 14.1 中的数据。

表 14.1 由三个取样单位内三个种(SPP)的丰富度组成的群落数据矩阵

种	SUs		
	(1)	(2)	(3)
(1)	20	15	0
(2)	10	0	6
(3)	17	0	0

表 14.2 计算取样单位(1)和(3)之间距离所需要的差值(DIF), 总和(SUM)及平方和(SSQ)。

种	SU		DIF		SUM (1+3)
	(1)	(3)	(1-3)	(1-3) ²	
(1)	20	0	20	400	20
(2)	10	6	4	16	16
(3)	17	0	17	289	17
	SUM=47	6	41	705	
	SSQ=789	36			

利用这个三个取样单位内三个种丰富度的简单数据矩阵,我们将说明取样单位(1)和(3)之间距离的计算方法。为了计算取样单位(1)和(3)之间的距离,必需有三种内和三种间的差值、总和及平方和(表 14.2)

距离 1. 欧氏距离 ED(14.1)

$$ED_{1,3} = \sqrt{[(20 - 0)^2 + (10 - 6)^2 + (17 - 0)^2]}$$
$$= \sqrt{400 + 16 + 289} = \sqrt{705} = 26.6$$

距离 2. 平方欧氏距离 SED(14.2)

$$SED_{1,3} = (400 + 16 + 289) = 705$$

距离 3. 平均欧氏距离 MED(14.3)

$$MED_{1,3} = \sqrt{(705/3)} = \sqrt{235} = 15.3$$

当比较这三个相互联系的方法时,我们可以注意到 SED 比 ED 和 MED 大得多,因为它是差值的平方和。取样单位(1)和(3)内差值最大的种在最终距离值上将得到最大的加权(例如,种 1 为 400, 种 2 为 16, 种 3 为 289)。

距离 4. 绝对距离 AD(14.4)

$$AD_{1,3} = |20 - 0| + |10 - 6| + |17 - 0| = 20 + 4 + 17 = 41$$

距离 5. 平均绝对距离 MAD(14.5)

$$MAD_{1,3}=41/3=13.7$$

当把这两个相互联系的方法和 ED, SED 及 MED 比较时, 我们可以发现丰富度差值较大的种的相对重要性较小, 这是因为它们的差值没有平方(例如种 1)。

距离 6. Bray-Curtis 相异性(14.6c)

$$DD_{1,3}=1-[(2)(0+6+0)/(47+6)]=1-(12/53)=0.77$$

距离 7. 相对欧氏距离 AED(14.7):

$$\begin{aligned} RED_{1,3} &= \sqrt{[(20/47)-(0/6)]^2 + \dots + [(17/47)-(0/6)]^2} \\ &= \sqrt{(0.426-0)^2 + \dots + (0.362-0)^2} \\ &= \sqrt{0.181+0.619+0.131} = \sqrt{0.931} = 0.96 \end{aligned}$$

距离 8. 相对绝对距离 RAD(14.9)

$$\begin{aligned} RAD_{1,3} &= |(20/47)-(0/6)| + \dots + |(17/47)-(0/6)| \\ &= |0.426-0| + \dots + |0.362-0| \\ &= (0.426+0.787+0.362) = 1.57 \end{aligned}$$

距离 9. 弦距离 CCOS。首先求出弦距离的余弦

$$\begin{aligned} CCOS_{1,3} &= [(20)(0) + (0)(6) + (17)(0)] / \sqrt{(789)(36)} \\ &= (0+6+0) / \sqrt{28404} = 60/168.5 = 0.356 \end{aligned}$$

那么弦距离(14.9a)

$$CRD_{1,3} = \sqrt{[12(1.0-0.356)]} = \sqrt{(2)(0.64)} = 1.13$$

距离 10. 测地学距离 GDD(14.10)

$$GDD_{1,3} = \arccos[0.356] = 1.21$$

14.4 距离函数的评价

在前一部分我们介绍了 10 种常见的距离函数。很明显, 这些函数中有一些彼此很相似, 而另外一些则有很大的区别。在这一部分我们将介绍一些怎样在不同的数据集上应用这些距离函数的示例。

表 14.3 基于表 14.1 给出的数据, 利用课文中讲的 10 种距离系数计算出的距离值

距离组	距离方法	公式	SU(j, k)		
			(1, 2)	(1, 3)	(2, 3)
E	ED	14.1	20.4	26.6	16.2
	SED	14.2	414.0	705.0	261.0
	MED	14.3	11.7	15.3	9.3
	AD	14.4	32.0	41	21
	MAD	14.5	10.7	13.6	7.0
BC	PD	14.6	0.52	0.77	1.00
RE	RED	14.7	0.71	0.96	1.41
	RAD	14.8	1.15	1.57	2.00
	CRD	14.9	0.76	1.14	1.41
	GDD	14.10	0.78	1.21	1.57

表 14.1 数据矩阵的所有可能取样单位对之间的距离都列在表 14.3 中。虽然取样单位 (2) 和 (3) 之间无共有种, 前五个计算距离的方法 (E 组) 实际表明这两个取样单位间的相似性大于取样单位 (1) 和 (2) 或 (1) 和 (3) 之间的相似性, 后两对之间都有一共有种。最后五个计算距离的方法 (BC 组和 RE 组) 没有得出这种不合理的结果。事实上, PD, RED, RAD, CRD 和 GDD 都给出相同的取样单位之间的距离顺序, 即取样单位 1 和 2 最相似, 取样单位 2 和 3 之间相似性最小, 这是一个更真实的结果。

从表 14.1 上, 我们第一眼凭直觉就可以感到似乎取样单位 1 和 2 之间的距离大于取样单位 1 和 3 之间的距离。

取样单位 1 和 2 的共有种的绝对差值为 5(种 1), 取样单位 1 和 3 的差值是 4(种 2)。然而利用相对权重计算这些指数得到的结果是取样单位 1 和 2 之间的距离小于 1 和 3 之间的距离。虽然表 14.1 中的数据集是人造的且很简单, 但它确实对说明当群落数据中出现 0 值(这是正常情况)时的一些难点有所补益, 且同学们应小心从事。

RED 和 RAD 表示的是相对于每一个取样单位总丰富度的种丰富度。RED 和 RAD 表达式的效应是均衡总丰富度高低不等的取样单位内种的重要性。各种丰富度比例大致相同的两个取样单位之间的相似性较大(距离较小)。因此, 如果你想要使总丰富度高的取样单位内高丰富度种的重要性和总丰富度低的取样单位内低丰富度种的重要性相似, 那么你应该采用 RED 和 RAD 方法。

弦距离(RD)和测地学距离(GDD)是比较相对于取样单位平方和丰富度的种丰富度。因此, 应用 RED 和 RAD 时, 当两个取样单位有大致相等的种丰富度比例时, 它们之间的距离较小。为了进一步说明这些相似函数的使用方法, 我们用表 11.4a 中的距离函数值。然后,

表 14.4 7 个取样单位内 11 种(A-K)的百分比丰富度数据, 对这些数据分别采用这部分中介绍的十种距离指数, 计算各取样单位之间的距离顺序关系

种	SUs						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A	100	0	50	100	5	0	85
B	90	10	50	40	0	0	0
C	80	20	50	20	5	0	65
D	70	30	50	10	0	5	0
E	60	40	50	5	5	0	75
F	50	50	50	0	0	5	0
G	40	60	50	5	5	5	0
H	30	70	50	10	0	0	65
I	20	80	50	20	5	0	85
J	10	90	50	40	0	5	0
K	0	100	50	100	5	5	0

利用 12.1 公式, 计算出所有取样单位对之间的十种距离函数(表 14.5)的 Spearman 秩相关系数。由 E 组(ED, SED, MED, AD 和 MAD)和 RE 组(RED, RAD, CRD 和 GDD)方法

计算得到的距离，在每一组内部高度相关；而这两组方法得到的距离之间相关很低。Bray—Curtis 的 PD 指数和 E—组计算结果之间相关较低，但它和 RE 组的计算结果有很好的相关 (0.56—0.67)。

表 14.5 十种计算距离的方法之间的 Spearman 秩相关。对角线上线下的相互关系是分别根据表 11.4a 和表 14.4 计算得出的

组		ED	SED	MED	AD	MAD	PD	RED	RAD	CRD	GDD
E	ED	1.0	0.99	0.99	0.99	0.99	0.27	-.25	-.24	-.22	-.22
	SED	1.0	1.0	0.99	0.99	0.99	0.27	-.25	-.24	-.22	-.22
	MED	1.0	1.0	1.0	0.99	0.99	0.27	-.25	-.24	-.22	-.22
	AD	0.86	0.86	0.86	1.0	1.0	0.27	-.26	-.24	-.21	-.21
	MAD	0.86	0.86	0.86	1.0	1.0	0.27	-.26	-.24	-.21	-.21
BC	PD	0.48	0.48	0.48	0.47	0.47	1.0	0.64	0.67	0.66	0.66
RE	RED	0.36	0.36	0.36	0.16	0.16	0.56	1.0	0.99	0.97	0.97
	RAD	0.31	0.31	0.31	0.20	0.20	0.63	0.92	1.0	0.97	0.96
	CRD	0.46	0.46	0.46	0.28	0.28	0.57	0.96	0.87	1.0	0.96
	GDD	0.46	0.46	0.46	0.28	0.28	0.57	0.96	0.87	1.0	1.0

从这些评价中，我们可以注意到以下几点：1. 尽管 E—组计算距离的方法应用很普及，但我们并不想推荐该组方法。从我们的计算结果(表 14.3)可以清楚地看到，它会出现不合逻辑的后果，Wolda(1981)也得到了相似的结论。

2. 每一个 RE—组函数看起来都很适用，而且结果合理。其中每种方法之间区别不大(表 14.5 给出了它们之间的高度相关)，但我们还是发现对于一系列不同的生态学数据，弦距离可能得到十分令人满意的结果。

3. PD 是 RE—组外的另一选择。Beals (1984)基于他广泛的生态学研究中 PD 的成功应用，对这个系数做了高度评价。

14.5 例子：巴拿马蟑螂

Basic 微机程序 SUDIST.BAS(见本书附带的软盘)是基于五种蟑螂的丰富度，计算六个巴拿马地点之间距离的(数据在表 11.4a 中)。其结果在表 14.6 中给出。请注意不同的距离方法它们的结果出入很大。例如，SED 值可高达数千，但也可低到 1.0。我们知道 SED 和其它 E—组方法(ED, MED, AD 和 MAD)变化范围是 0 到无穷大，这是因为他们随着种数的增加而增加。相反，我们知道 RED 和 CRD 有一个上限—— $\sqrt{2}=1.41$ ，RAD 有一上限——2，GDD 也有一上限—— $\pi/2=1.57$ 。

表 14.6 基于 5 个种的丰富度, 应用 SUDIST.BAS 计算巴拿马六个地点之间的距离大小。

距离方法											
地点		E—组					BC—组		RE—组		
(j)	(k)	ED	MED	SED	AD	MAD	PD	RED	RAD	CRD	GDD
1	2	52.2	23.3	2722	94	18.8	0.50	0.49	0.83	0.74	0.76
1	3	74.6	33.4	5571	115	23.0	0.97	0.74	1.28	1.10	1.16
1	4	75.0	33.5	5626	116	23.2	0.98	1.08	1.76	1.28	1.38
1	5	64.2	28.7	4127	101	20.2	0.76	0.25	0.47	0.34	0.34
1	6	54.5	24.4	2966	94	18.8	0.66	0.47	0.84	0.44	0.44
2	3	47.1	21.1	2219	69	13.8	0.95	0.60	0.89	0.87	0.90
2	4	47.2	21.1	2226	70	14.0	0.97	0.62	0.93	0.64	0.66
2	5	38.6	17.3	1489	55	11.0	0.63	0.40	0.57	0.56	0.57
2	6	38.5	17.2	1486	48	9.6	0.50	0.77	1.18	0.89	0.92
3	4	1.0	0.4	1	1	0.2	0.33	0.71	1.00	0.77	0.79
3	5	11.4	5.1	130	14	2.8	0.78	0.85	1.38	1.18	1.27
3	6	24.1	10.8	579		5.4	1.00	1.19	2.00	1.41	1.56
4	5	11.4	5.1	131	15	3.0	0.88	1.02	1.50	1.15	1.22
4	6	24.0	10.8	578	26	5.2	1.00	1.39	2.00	1.41	1.56
5	6	13.7	6.1	187	19	3.8	0.46	0.38	0.63	0.36	0.36

14.6 示例: 威斯康星森林

根据威斯康星南部 10 个高地森林地点 8 种树的数据矩阵(表 11. 6a), 利用 SUDIST.BAS 程序(见附带的软盘)计算 10 个地点中所有两两地点之间的弦距离(CRD)(表 14. 7)。我们知道 CRD 的最大值为 1. 41(最大相异性), 很显然, 取样单位(地点)1 和 2 之间最相似, 其次为取样单位(地点)9 和 10。

表 14.7 SU×SU 矩阵形式(对角线以上)的威斯康星南部 10 个高地森林地点之间的弦距离(CRD)。

SUs	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	0.15	0.61	0.50	0.83	1.17	1.02	1.22	1.35	1.30
(2)		0.64	0.50	0.85	1.16	1.02	1.22	1.33	1.29
(3)			0.45	0.94	1.14	1.12	1.18	1.38	1.25
(4)				0.57	0.95	0.90	1.05	1.28	1.18
(5)					0.79	0.67	0.95	1.16	1.07
(6)						0.74	0.41	0.80	0.80
(7)							0.63	0.62	0.61
(8)								0.52	0.53
(9)									0.31

通常,这些结果用于进一步的分析,如聚类分析(第十六章)。为了这样的分析,距离为 $SU \times SU$ 比较矩阵的形式用起来比较方便(象表 14.7 列举的 CRD 距离的 10×10 矩阵的示例。)

14.7 相似性函数的附加讨论

Hubalek (1982)“根据五个主要条件对 43 种用于有一无数据的 Q-方式相似函数的相似系数的可采用性做了评价。”Hubalek 提出对于一个试验数据集仅有四个相似系数使用效果较好:(1)Jaccard 的群落系数,(2)Dice 一致性指数,(3)Kulczynski 系数,(4)Ochiai 系数。对 Jaccard, Dice 和 Ochiai 系数,在 Janson 和 Vegelius (1981)的一篇关于 20 种相似性方法的评论文章中也有高度的评价。

Wolda (1981)根据信息量,利用 22 种生态学相似方法(包括积矩和顺序相关系数),对样本的大小和种多样性的效应做了检测。Wolda 也没有推荐把相关系数作为相似性指数。在他检测的信息方法中,Wolda 对 Morisita (1959)指数做了高度评价,因为他证明了该指数不受取样大小和多样性的影响。当数据需要事先进行对数变换时,Wolda 推荐使用经 Horn (1966)简化后的 Morisita 指数和 Horn 重叠指数。然而,Bloom (1981)发现,Horn 的这两个指数“彼此之间以及和理论标准之间都有很大的分歧”(该标准基于一个常规曲线范围表)。

生态学家还没有大量使用相似性的概率方法,例如,Goodall (1964, 1966)和 Feoli 与 Lagonero (1983)提出的指数。这部分原因是由于它们较复杂以及较冗长的计算。Goodall 指数的主要优点是:(1)它的范围是 0—1,(2)它是线性的,(3)它即可用于丰富度数据,也可用于有一无数据(Orloci 1978)。其主要缺点是任何两个取样单位之间的相似性概率是基于数据集中所有取样单位的,因此,如果从一个数据集中加上或减去取样单位,那么给定的两个取样单位之间,基于概率的相似性就要发生变化。

生态学家对信息指数用得也很少,例如,Horn 的重叠指数—用于计算取样单位之间的相似性。Orloci (1978)提出了另一些取样单位之间相似性的信息方法。在不同相似性函数的优秀评论文章中,我们向同学们推荐 Boesch (1977),Orloci (1972, 1978)Campbell (1978),Clifford 和 Williams (1976),Goodall (1978b),Hubalek (1982),Orloci (1972, 1978),Pielou (1984)和 Williams 与 Dale (1965)的文章。

许多生态学数据集是取样单位内种的密度、频度、盖度、生物量等数量信息的混合。有些种可能在一些取样单位中占优势,而在另一些取样单位中不存在。也可能有的种在全部样本中很少见。为了避免在数据分析中过度强调优势种,生态学家在计算生态相似性之前,经常需要对数据进行标准化或进行变换(Jensen 1978)。大量的变换类型是可以使用的,其中有一些我们已在 14.2.2.3 中做过介绍。Chardy 等(1976)在分析仅几种占优势的高多样性浮游生物群落之前使用了对数变换。其它变换也受到赏识(如,角变换和平方根变换)。Gauch (1982),Greig-Smith (1983),Jensen (1978),Noy-Meir (1973),Noy-Meir 等(1975)和 Orloci (1978)为我们提供了一个极好的生态学数据变换的相对优点的总看法。Hajdu (1981)在用图解法比较 16 种相似性方法时发现,用取样单位进行标准化,在排序系列上会产生不良影响。

当然,寻找新的(相对于早期提出的函数)生态学相似性函数仍会继续下去。例如,基于一元(二元数据)数据的新相似性函数(Faith 1983, 1984)以及计算距离的函数(Bradfield 和 Kenkel 1987)都已被提出,并进行了试验。其真正的生态学价值将来自它们被应用于分类和排序过程后,对最有生命力的方法的认识。

14.8 总结和推荐

1. Q-方式(SU)相似性系数可以分成两个类型:相似性系数和距离系数(14.1 部分)。
2. 当数据是种的有一元数据时,我们推荐用 Ochiai, Dice 和 Jaccard 相似性系数计算取样单位之间的相似性(14.2.1 部分)。
3. 基于丰富度数据的距离函数可分成三类:E-组(欧氏距离指数),BC-组(以 Bray-Curtis 相异指数为代表)和 RE-组(相对欧氏距离指数)(14.2.2 部分)。
4. 尽管 E-组距离指数应用很广,但我们不推荐使用它。虽然这些指数在生态学上有很大的启发意义,但在使用时会出现各种错误。
5. 当数据是数量化的丰富度数据时,我们推荐使用 RE-组距离函数中的弦距离计算取样单位之间的相似性。

第十五章 关联分析

在这一章中我们说明一下,作为分类技术的关联分析的应用。取样单位(SUs),例如,样方和样地,根据划分规则按等级地分成同质的几组。当种间关联(比如由有一无数据决定的)消失时,便认为该组是同质的。

15.1 基本方法

关联分析(Association analysis, AA)起源于 David Goodall (1953)开创性的工作,也许因为这种方法的直观性,使它成为第一个被用于植物生态学分类的方法。AA 技术的基础假设是由 Goodall 阐述的,即“如果一个样方数据集表明,样方内两个种的数量是关联的,那么这些种中的一或两种在这个地区的分布可能是不均匀的,而且控制它们的生境条件也是不均匀的。”因此,在 AA 分析中可以表明,有些种类对控制群落结构的环境因子比其它种类更敏感,因此那些(敏感)种可以形成分类的基础(Coetzee 和 Werger 1975)。

以下介绍的 AA 方法选自 Williams 和 Lambert (1959)。给出一个取样单位集合,如果有种和其它种表现出显著关联(基于计算所有可能 2×2 表卡方值第十一章),那么我们认为这个集合是“异质的”。异质性可以被取样单位集合的进一步划分减小,因此显著的种间关联将消失(或者关联的数量会被减少)。因此反复地进行划分,形成的各组内部的异质性就会减小,而各组之间的异质性就会增加。当形成的一组内种间关联消失时,我们称该组是“同质的”。最后的结果是基于种间关联上的取样单位的分类。关联分析被成功地应用于大量各种各样的植被类型,从森林到荒漠(见表 13.1)。

15.2 步骤

关联分析是一种单元的、分裂的、等级的技术,正象第十三章中定义的那样。AA 的目的是对取样单位进行分类,因此认为 AA 是一种 Q-方式分析,虽然该过程仅基于 SU 内种的有一无数据,而且用的是种间关联的成对比较。

15.2.1 计算步骤

第一步:形成 2×2 列联表。取样单位集合内每一个可能结合的种对的 2×2 列联表都是在有一无数据基础上的形成的(见第十一章)。对于 S 个种,则有 $(S)(S-1)/2$ 个种对。

第二步:计算 χ^2 统计量。每个列联表,可以计算出一个 χ^2 值(11.7 部分)。这些值和临界值表中 χ^2 值 3.84 ($P=0.05$, $df=1$) 相比较以便确定其显著度。如果 $(S)(S-1)/2$ 个 χ^2 值中没有一个达到显著程度,那么可以认为取样单位集合是同质的,过程便可终止了。另一方面,如果任何两种间存在显著关联,则认为取样单位集合是异质的,第三步则应该继续进行。

许多研究者已指出临界 χ^2 值——3.84 不适于同时 χ^2 检验,需要一个更大的临界值(见

Jensen 等, 1968)。我们在 11.2.3 部分中已讨论了这个问题, 并介绍了 Schluter (1984) 用于检验显著相关的方差比率检验。我们将在 15.2.2 部分再次谈及此论题。

第三步: 显著 χ^2 值的加和。计算出每个种达到显著程度的 χ^2 值的 (>3.84) 总和。如果一个种没有显著相关, 则它的总和为零。然而, 例如, 如果一种和其它四种显著相关, 那么它的总和应仅根据这四个值计算。

第四步: 选择区分种。总显著 χ^2 值最高的种就是“区分”种。

第五步: 对取样单位进行分类。根据取样单位内区分种的有无, 把取样单位分成两组。这个结果是原始取样单位组再分为两组, 一个组包含有选择出的区分种, 另一组中则没有区分种。

如果区分出的两组内部已经是同质的, 为了检验必需为每个新组建立一新的 χ^2 值集合(第一步)。如果没有一个取样组内有种的显著相关, 那么我们的目的就达到了, 每一个组都可以认为是同质的取样单位的集合。然而, 如果一组或两组取样单位中有显著的种间相关, 第 1—5 步就应该继续下去, 直到再没有显著相关存在为止, 即这个过程重复下去, 直至所有组内都达到同质性。AA 的结果是取样单位组的形成, 每一组内的同质性都大于各组之间的同质性。

AA 的主要问题是: 是否和什么时候可以达到同质性。下面将提出并讨论停止规则。

15.2.2 停止规则

Goodall (1953) 的原始 AA 方法是连续划分取样单位组, 直到显著 χ^2 值 (>3.84) 不再存在。然而, 这种方法常常形成许多小的取样单位组。为了克服这一点, Goodall 对最终组进行了各种联合, 并依次再次检验了它们的同质性。联合组中如果不产生任何的显著种间关联, 我们就把联合组为一个 SU 的最终组。Goodall 的联合过程的主要结果是否定了关联分析的等级性。

Williams 和 Lambert (1959) 选用在某一水平停止进一步划分的方法, 来完全保持 AA 的等级性, 超过该水平则分类结果会过分破碎。他们推理, 划分应停止于某一水平, 该水平和被检测组内的 SU 数量成比例, 而不是在没有显著种间关联存在时才停止划分。他们根据这一原则, 提出了几个划分的数学标准。

Madgwick 和 Desrochers (1972) 提出, 停止水平应和种数 (S) 成比例, 而不是取样单位数 (N)。他们认为是关联种对数 $[(S)(S-1)/2]$ 导致卡方值 -3.84 成为不适当的显著度水平 (因为 3.84 是单对检验的 5% 的显著度水平)。他们推荐人们使用由 Jensen 等 (1968) 制做的列联表中的同时卡方检验数目 (种对结合) 决定的显著度水平。例如, 20 个种对之间的关联同时比较中, 9.885 这个卡方值是适当的 ($P=5\%$)。

我们更喜欢用“试—错”停止规则。我们采用的是 Goodall 的分组标准, 即直到不再存在显著的卡方值 (>3.84) 为止, 但在最后的分析中, 划分的停止水平是根据在什么水平上才能产生最有生态学意义的取样单位组而定的。利用这个规则, 3.84 这个卡方值只能被看作是一个方便的指数, 而不是一个显著度的统计检验。虽然这个“试错”规则看起来不如前面的其它临界值系数那样客观, 但它确实能使我们避免造成取样单位组被划分的过于破碎。同时对仅有几个取样单位的组也不应该再进行细分。在一篇关于 AA 的评论文章中, Madgwick 和 Desrochers

(1972)提出,在考虑划分之前,每个取样单位组内应至少有 6 个取样单位。

15.3 举例:计算

利用表 11. 4b 中巴拿马蟑螂的数据(6 个地点,5 个种),AA 的过程如下:

第一步和第二步:显著种间关联的检验结果在表 15. 1 中给出,只有种 1 和种 2 之间的卡方值大于 3. 84($X^2=6. 0$)(虽然由于样本太小,会造成这个值有偏差,但我们还要继续这个示例)。种 1 和种 2 从来不一起出现于任何取样地点,它们表现出负相关(表 11. 4b)。

表 15. 1 六个地点 5 个种的巴拿马蟑螂种对的种间关联显著度检验的卡方值。应用的
是计算机程序 SPASSOC. BAS。

种对	关联类型 ^a	X ²	
		值	偏差
1 2	—	6.00	*
1 3	—	2.40	*
1 4	—	0.24	*
1 5	+	0.60	*
2 3	+	2.40	*
2 4	+	0.24	*
2 5	—	0.60	*
3 4	+	0.60	*
3 5	+	0.38	*
4 5	+	0.60	*

a: 表示种间关联的方向

*: 因为 2×2 表任一格内的期望频度小于 1 和/或有两个以上格内的期望频度小于 5, 所以未经校正的卡方值是有偏差的。

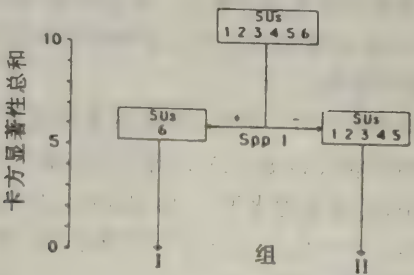


图 15. 1 基于 5 种蟑螂有无数据,利用关联分析划分六个巴拿马地点取样单位

第三步:显著卡方值之和。从表 15. 1 可以看出种 1 和 2 的显著卡方值之和等于 6. 0,而种 3,4,5 不存在显著关联。

第四步:选择区分种。种 1 和种 2 具有最高的显著卡方值之和。在这种情况下,我们选择两者中任意一个作为区分种(对于这个示例,不论是种 1,还是种 2 作区分种都是一样的)。

第五步：根据区分种的有无划分取样单位。取样单位被分成含有种 1 的 I 组(SU₆)和不包含种 1 的 II 组(SU₁₋₅)图(15.1)。在每一组内再分别计算种间关联,以检测它的同质性。因为 I 组的所有取样单位内都没有种 1,种 1 就不再参与新的计算。因为 II 组仅有 5 个取样单位,进一步的划分就没有必要了。

15.4 例子:威斯康星森林

威斯康星森林数据(8 个种,10 个地点)的关联分析是利用 BASIC 程序—NASSOC.BAS (见附带的软盘)处理的,处理过程是按 15.2 部分中介绍的步骤进行的。

在 11.5 部分中我们接受了 8 种树之间没有关联的假设。这是基于同时种对关联方差比率检验。这些结果表明:AA 并不是处理这些数据的适当方法,因为并非全部种对之间都存在关联。然而,表 11.7 中给出的卡方值表明许多种对间存在显著的关联(因为样本太小都有偏差)。根据我们选定的同质性临界值,我们利用 AA 进行处理。

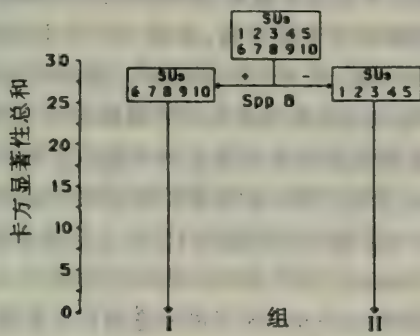


图 15.2 基于 8 个树种的有一无数据,利用 AA 对威斯康星南部高地森林的 10 个地点(SUs)进行划分

在图 15.2 中给出的 AA 结果,表明 10 个取样单位的集合并不是同质的。种 8 具有最大的显著卡方值之和(26.7—同学们在表 11.7 中可以很快查到这个结果)。根据种 8 的有无,将取样单位划分成两组,每组中包括 5 个取样单位(I 组包括取样单位 6,7,8,9,10, II 组中包括取样单位 1,2,3,4,5)。因为每组中取样单位都少于 6 个,因此 I 组和 II 组不再继续划分。

如果我们现在再次检测威斯康星森林的数据,我们可以看到这些结果有一定的生态学意义。代表 I 组的五个取样单位(6—10SUs),以顶极森林种为特征,它们是糖槭(区分种)、椴树和硬木,而 II 组中的取样单位(SU₁₋₅)倾向于以演替森林种为特征,它们是栎树(Peet 和 Loucks 1977)。虽然我们在整本书中一直提醒同学们不要使用小的数据集,但这个示例帮助我们说明,AA 仅是一简单的分类模型。

15.5 关联分析的附加讨论

在这一章中我们主要介绍了叫做一般关联分析(Normal association analysis)AA 的一个类型,它是根据种的有无进行分类的。只要简单地转化一下表格(变换数据矩阵),用同样的 AA

进行处理,我们便可以形成种的分组。在生态学文献中这种类型的AA被叫作逆关联分析(Inverse association analysis)(Williams 和 Lambort 1961)。具有生境(SU)发生相似性的种将被分在同一组内。此外,一般关联分析和逆关联分析的结果可以制成交叉表,以便寻找取样单位和种组之间的偶合结点(结点分析—Boesch 1977, Lambert 和 Williams 1962, Noy—Meir 1971)。

Madgwick 和 Desrochers (1972)建议在AA中应该排除稀有种。他们的研究表明,在AA分析中如果包括稀有种将会增加 2×2 列联表中每一格内频度小于3的机会,这会造成分组过于破碎的后果。他们建议用校正卡方值(虽然他们发现Yates的校正有些“过头”)或Fisher的精确检验。Pielou (1977)对此也提出过建议。采用Yates校正因子的观点也可用于我们的程序NASSOC.BAS,而且我们鼓励同学们检验一下,在AA中选用可选择卡方值的效果。

利用39种和167个取样单位(森林样地),Madgwick 和 Desrochers (1972)揭示了所有单元分类技术共有的一个问题:即由于区分种的机率发生(事件),造成取样单位的错误分类。因为AA技术基于种的有一无数据,一个取样单位内(事实上,它可能是以其它种占优势)区分种的机率事件可以导致不合逻辑的和谬误的结果。利用判别分析检验他的AA分类(见第二十三章),他们发现由于正选择即根据区分种的存在,造成167个SU(样地)中的27个SU被错误分类,而根据缺乏区分种进行的分类,仅有五个样地(SU)被错误地分类。

单元、分裂分类对策较大的优点就是当涉及到大量取样单位时,例如,大规模生态学调查中,该方法的计算效率很高。从所有取样单位都在一个组内开始,不断划分直到获得满意的同质取样单位组,这种方法比开始于每个独立的取样单位而逐步聚合(聚合对策,第十六章)的方法在计算上更有效。这里还介绍了其它单元划分技术,包括划分信息分析(Lance 和 Williams 1968)和组分析(Crawford 和 Wishart 1967、1968),但以上介绍的AA简单过程仍是一种基础方法。多元划分技术也被提出(Greig—Smith 1983),它减少了单元分析方法的分类错误。

15.6 总结和推荐

1. 关联分析(AA)是一种基于减少一个取样单位组内的异质性的分类方案。如果一组内存在显著的种间关联,我们则认为它是异质的(基于所有种对的 2×2 表,第十一章)。利用划分减小取样单位集合的异质性,以便使显著的种间关联消失。

2. 为了了解显著关联的存在,所有种对的 2×2 表都要进行检测,所以我们推荐在进行AA之前,先使用方差比率检验(第十一章)。如果同时检验表明数据中不存在显著种间关联,那么可以使用AA。同学们应该始终认识到这样一个事实,即象AA这样的分析方法可以产生一个分类,但其结果可能是偶然的。

3. 作为AA中的停止规则我们推荐“试错”法和Goodall划分取样单位组直到没有卡方值大于临界值(3.84)为止的方法相结合,也就是说分类应停止在得到最有生态学意义的分类时。这儿并未指出统计学意义。

4. 我们还推荐,一个组内取样单位数的最小值为6,即当一个组内少于6个取样单位时,则不再继续划分。

第十六章 聚类分析

聚类分析(Cluster Analysis, CA)是一种把相似的实体放入组或“聚类”内的分类技术。我们在这章中将介绍的聚类分析模式是把相似的样本放入聚类内,它是一个有等级的树状结构,叫做树状图。这些取样单位的组或聚类可以划分成或代表不同的生物群落。

16.1 基本方法

给出一个实体集合和一些它们相互之间相似性的计算方法,我们定义把这些实体“归入”组或聚类内即为分类。聚类分析是完成这种归类的一种技术。这儿涉及到的实体是生态学样本或取样单位(例如,样地、样条、样方)。CA 实际上是一个总括性术语,它涉及到大量的算法规则,它们之间的区别主要是在聚类形式的处理上。

在详述具体过程之前,我们首先介绍 CA 的一般方法。开始,我们必须计算取样单位之间的 Q-方式相似性。虽然有许多相似函数可以利用,但我们仅限于距离方法(见第十四章),因为它们在 CA 中有启发意义(Sneath 和 Sokal 1973)。集合内所有取样单位对之间的距离都概括在 $SU \times SU$ 距离或 D 矩阵内(例如,表 14.7),并对此进行各种 CA 对策处理。

这章中描述的 CA 模式是聚合的(第十三章):它们开始于一个独立取样单位的集合,并逐步建立相似取样单位的组或聚类。在每一次聚类中仅能有一对实体相互结合形成一个新的聚类。这对实体可以是:(1)一个取样单位和另一取样单位,(2)一个取样单位和一个已存在的聚类,(3)一个聚类和另一聚类。因此使用了组对 CA 这个术语。

组对方法的一个一般示例在图 16.1 中进行了说明。因为这个示例中我们用了 5 个取样单位,因此 10 个种对 $[N(N-1)/2]$ 之间的距离形成了 D 矩阵。所有组对 CA 对策的第一步都是在 D 矩阵中寻找两个取样单位之间的距离最小值。在图 16.1 中,这个最小值是取样单位 1 和 4 之间的距离,用符号 j 和 k 表示。因此,第一个聚类是在距离 $D(j,k)$ 处形成,这个结果可用一个树状图画出(图 16.1,第一轮)。原始的五个取样单位的集合变成了一个聚类($C_1 = SU1 + SU4$)和 3 个独立的取样单位(2,3,5)。现在必须计算这个聚类和三个剩下的取样单位之间的距离。

用于这种类型计算的特殊方程已被提出, Lance 和 Williams (1967)提出的一个通用方程——线性结合方程,在下面给出。

线性结合方程形式如下:

$$D(j,k)(h) = \alpha_1 D(j,h) + \alpha_2 D(k,h) + \beta D(j,k) \quad (16.1)$$

其中,来自第 j,k 取样单位的新聚类(j,k)和第 h 个取样单位或取样单位组之间的距离可以从已知的距离 $D(j,k)$, $D(j,h)$ 和 $D(k,h)$ 及参数 $\alpha_1, \alpha_2, \beta$ 计算得到。例如,取样单位 3 和聚类 ($SU1 + SU4$) 之间的距离(图 16.1,第 2 轮),可用下式计算:

$$D(1,4)(3) = \alpha_1 D(1,3) + \alpha_2 D(4,3) + \beta D(1,4) \quad (16.2)$$

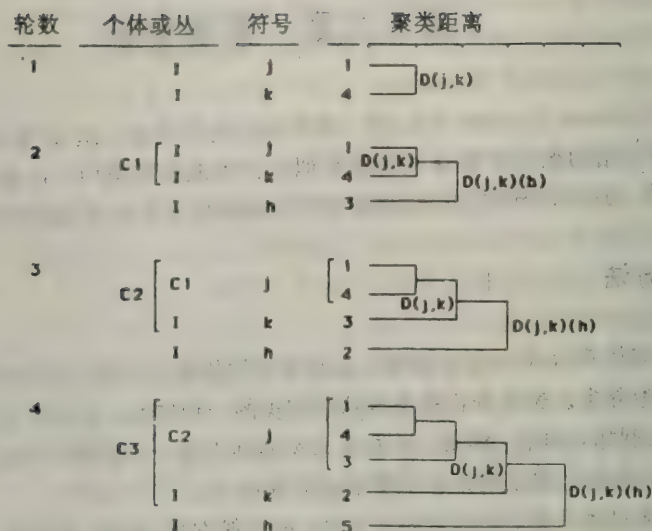


图 16.1 Lance 和 Williams (1967)对于 5 个取样单位(见方程 16.1)的结合聚类方法的说明

表 16.1 不同的等级聚类对策中, Lance 和 Williams 结合方程(16.1 部分)中参数值 $\alpha_1, \alpha_2, \beta$ 。
对策的名称是根据 Sneath 和 Sokal 1973)/Lance 和 Williams (1967)第 j 和 k 组内
取样单位数是 $t(j)$ 和 $t(k)$, 结合组 (j,k) 内取样单位数是 $t(j,k)$

对策	α_1	α_2	β
矩心(不加权)/矩心	$t(j)/t(j,k)$	$t(k)/t(j,k)$	$-t(j)t(k)/t(j,k)$
矩心(加权)/中数	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{4}$
组平均(不加权)/组对方法	$t(j)/t(j,k)$	$t(k)/t(j,k)$	0
可伸缩	0.625	0.625	-0.25 ^a

^a 在该对策内 β 是可伸缩的, 其中 $\alpha_1 + \alpha_2 + \beta = 1$, $\alpha_1 = \alpha_2$ 。

不同的聚类对策, 只是它们的 $\alpha_1, \alpha_2, \beta$ (表 16.1)的值不同, 它们($\alpha_1, \alpha_2, \beta$)是决定新距离的加权(在下面进一步加以说明)。

方程 16.1 中给出的取样单位间的关系和新聚类的形式, 在图 16.1 中进行了描述(j, k, h 的关系在 16.1 方程中表明了), 从图 16.1 上可以看出:

1. 在一个有 N 个取样单位的集合中, CA 中包括 $N-1$ 轮。在该示例中有 4 轮。
2. 在第一轮中两个取样单位(用 1 表示)结合形成一个聚类。取样单位 1(符号 j)和 4(符号 k)结合的距离是 $D(j,k)$, 该值来自 D 矩阵。
3. 在第二轮中, 取样单位 3(符号 h)和第一轮形成的聚类(符号 C_1)结合。 j 和 k 分别代表取样单位 1 和 4, 取样单位 3 和 C_1 结合的距离是 $D(j,k)(h)$ 。
4. 第三轮, 取样单位 2(符号 h)和在第二轮中形成的聚类(符号(C_2))结合。注意现在 j 代表

的是聚类 C_1, k 是和 C_1 结合的最后一个取样单位。

5. 在第四轮, 取样单位 5(符号 h) 和聚类 C_3 结合。在 16.1 方程中, j 是聚类 C_2 ($SUs 1, 3, 4$), k 是 $SU 2$ 。

正如前面提到的, 方程 16.1 中的 α 和 β 是决定距离的“加权”。聚类结果会因为使用的加权方案不同而变化。在一些情况中, 差异是巨大的。表 16.1 中给出了四种特定的加权方案; 矩心(Centroid)(加权和不加权)对策、组平均(group-average)对策及可伸缩的(Flexible)对策。

加权的概念可能最好是用一示例来说明。我们再看一下图 16.1 中的示例, 聚类 C_3 (此处 $j = SUs 1, 4$ 和 $3, k = SU 2$) 和 $SU 5$ 之间距离的组平均加权由下式给出(16.1 方程)。

$$D(1, 4, 3; 2)(5) = \frac{3}{4}D(1, 4, 3; 5) + \frac{1}{4}D(2, 5) \quad (16.3)$$

其中: $\alpha_1 = \frac{3}{4}$, $\alpha_2 = \frac{1}{4}$, $\beta = 0$ (来自表 16.1)。

组平均聚类对策(算术平均不加权组对方法)有效地计算了一组中取样单位和另一组取样单位之间所有距离的平均值, 因此是不加权的。在另一方面, 矩心加权对策(表 16.1)的结合方程是

$$D(1, 4, 3; 2) = \frac{1}{2}D(1, 4, 3; 5) + \frac{1}{2}D(2, 5) - \frac{1}{4}D(1, 4, 3; 2) \quad (16.4)$$

它不考虑每组内取样单位数的不同, 而对所有合并组都同等看待。我们也注意到, 一旦一个组形成, 它将被其平均值代替, 聚类之间的距离便是这些平均值或矩心之间的距离。

16.2 步骤

各种各样的聚类过程是在所有可能取样单位对的距离矩阵上进行的(表 14.7)。我们可以利用任何一个在第十四章介绍过的距离方法。我们假设在集合中共有 N 个 SU 。

第一步: 获取初始组。我们在 $N \times N$ 的 D 矩阵内寻找取样单位对之间的最小距离。这一对代表的是集合中最相似的两个取样单位。这两个取样单位相互结合(例如, 图 16.1)。

第二步: D 矩阵的缩小,现在集合中有 $N-1$ 个实体, 换句话说, 该集合由一个组(两个取样单位组成), 以及剩下 $N-2$ 个取样单位组成。新形成的组和剩下这些取样单位之间的距离可由方程 16.1 计算得到, 形成一缩小的新的 D' 矩阵 $(N-1) \times (N-1)$ 。

第三步: 搜索缩小的 D' 矩阵。正象在第一步中, 在新的 D' 矩阵中寻找最小距离值, 以便形成下一个新组。

第四步: 重复第二、三步,直到所有取样单位都结合成一组。这将需要进行总共 $N-1$ 轮, 因为任何的一轮计算仅能对一对实体($SU-SU$ 、聚类- SU 或聚类-聚类)进行聚类结合。任何一轮开始时(第二步)的实体数是 $N-C$, 其中 C 是轮数。

CA 的最后一个问题是, 一旦聚类完成, 特别组或群体的识别。象图 16.1 中显示的树状图, 可以用于检测取样单位的分组。然而, 这在很大程度上是一个主观决定, 最近已有一些试图使这些决定表现出更大的客观性的试验(例如, Hill 1980, Popma 等 1983, Ratliff 和 Pieper 1981, Rohlf 1974, 1982)。这些客观过程将在 16.6 部分中讨论。总的指导方针是不要使最后的划分结果成为大量过于破碎和不能解释的组。

16.3 举例:计算

Lance 和 Williams (1967)的结合线性模式[16.1 方程]在应用欧氏距离 D 矩阵(表 16.2)时已作了说明。这些距离是通过计算 5 个取样单位(表 11,3a)中三个种的丰富度的人造数据得到的。每次聚类合并后 D 矩阵的再建在表 16.2 中也给出了。而且,对可伸缩 CA 对策也作了说明。

第一步:获得初始组。D 矩阵中最小的欧氏距离是取样单位 2 和 3 之间的 1.41。因此,这两个取样单位形成第一组,这可以画在象图 16.2 中显示的那样的树状图中,这是第一轮聚类。

第二步:缩小 D 矩阵。这个新组(2,3)和剩下三个取样单位间的距离用 16.1 方程中等式计算:

$$\begin{aligned} D(2,3)(1) &= 0.625 \times 4.69 + 0.625 \times 5.10 - 0.25 \times 1.41 \\ &= 2.93 + 3.19 - 0.35 = 5.77 \\ D(2,3)(4) &= 0.625 \times 2.24 + 0.625 \times 3.00 - 0.25 \times 1.41 \\ &= 1.40 + 1.88 - 0.35 = 2.93 \end{aligned}$$

表 16.2 基于表 11.3a 中数据,5 个取样单位之间欧氏距离的 D 矩阵。只显示了右上角:
(a)初始 D 矩阵,(b)–(d)取样单位合并后的缩小 D 矩阵

		取样单位(SU)			
(a)		(2)	(3)	(4)	(5)
D=	(1)	4.69	5.10	3.00	2.24
	(2)		1.41	2.24	5.74
	(3)			3.00	5.92
	(4)				3.74
(b)			(2,3)	(4)	(5)
	(1)		5.77	3.00	2.24
	D'=(2,3)			2.93	6.94
	(4)				3.74
(c)				(2,3)	(4)
			D''=(1,5)	7.39	3.66
			(2,3)		2.93
(d)					(2,3,4)
				D'''=(1,5)	6.18

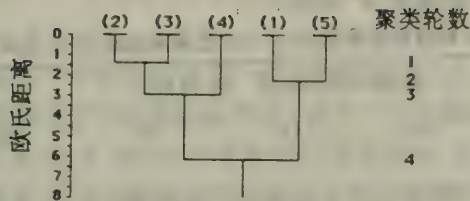


图 16.2 利用欧氏距离和可伸缩对策对五个取样单位进行聚类的树状图。(β=-0.25)

$$\begin{aligned} D(2,3)(5) &= 0.625 \times 5.74 + 0.625 \times 5.92 - 0.25 \times 1.41 \\ &= 3.95 + 3.70 - 0.25 = 6.94 \end{aligned}$$

缩小的 D 矩阵显示在表 16.2b 中。注意未进行聚类的取样单位之间的距离保持不变。

第三步: 搜索缩小的 D' 矩阵。D' 中的最小距离是取样单位 1 和 5 之间的 2.24。因此, 这两个取样单位形成图 16.2 中的第二个聚类, 这是第二轮聚类。

第四步: D' 矩阵的缩小。这个新聚类和剩下的取样单位 4 及组 (2,3) 之间的距离, 按下式计算:

$$\begin{aligned} D''(1,5)(2,3) &= 0.625 \times 5.77 + 0.625 \times 6.94 - 0.25 \times 2.24 \\ &= 3.61 + 4.34 - 0.56 = 7.39 \end{aligned}$$

$$\begin{aligned} D''(1,5)(4) &= 0.625 \times 3.00 + 0.625 \times 3.74 - 0.25 \times 2.24 \\ &= 1.88 + 2.34 - 0.56 = 3.66 \end{aligned}$$

注意: 来自前一轮的缩小 D' 矩阵被用于计算获取所有的新距离。这下一个新缩小的 D'' 矩阵显示在表 16.2c 中。

第五步: 搜索这个 D'' 缩小矩阵。D'' 中最小的距离是组 (2,3) 和取样单位 4 之间的 2.93。因此这三个取样单位在距离为 2.93 处形成了一个新聚类—正象图 16.2 中显示的那样, 这是第三轮聚类。

第六步: D'' 矩阵的缩小。包含三个取样单位的这个新聚类和由取样单位 1 和 5 组成的组之间的距离计算如下:

$$\begin{aligned} D''' = (2,3,4)(1,5) &= 0.625 \times 7.39 + 0.625 \times 3.66 - 0.25 \times 2.93 \\ &= 4.62 + 2.29 - 0.73 = 6.18 \end{aligned}$$

第七步: 最后的缩小矩阵显示在表 16.2d 中, 最后的合并并在欧氏距离等于 6.18 处, 将所有取样单位归到一起。这在图 16.2 中作了说明, 这是第四轮聚类。

16.4 例子: 巴拿马蟑螂

利用 BASIC 程序 CLUSTER.BAS (见附带软盘) 计算了 (表 11.4a) 中巴拿马蟑螂的数据集的 CA。在六个地点有五种蟑螂, 采用的是相对欧氏距离 (RED, 见 14.2.2.3 部分) 作为相似函数, 并且选用可伸缩的聚类对策 (见表 16.1)。Basic 程序的输出结果的总结在表 16.3 中给出。注意一个聚类是以最低数值 (SU 序号) 的 SU 序号来命名的 (例如, 在第二轮中的聚类

由 SU1,5 和 6 组成, 我们则称该聚类为“聚类 1”)。

表 16.3 程序 CLUSTER.BAS 的分析结果给出(a)六个巴拿马地点(SU)之间的距离和(b)地点的聚类

(a)相对欧氏距离 (D 矩阵)					
SUs	(2)	(3)	(4)	(5)	(6)
(1)	0.49	0.74	1.08	0.25	0.47
(2)		0.60	0.62	0.40	0.77
D=(3)			0.71	0.85	1.19
(4)				1.02	1.39
(5)					0.35

(b)用可伸缩对策进行聚类: $\beta = -0.25$				
聚类轮数	组序	聚类水平	参考 SU*	组中的 SUs
1	5	0.25	1	5
2	4	0.47	1	5, 6
3	3	0.60	2	3
4	2	0.68	2	3, 4
5	1	1.43	1	来自一组的所有 SUs

a 组中取样单位的最小数值

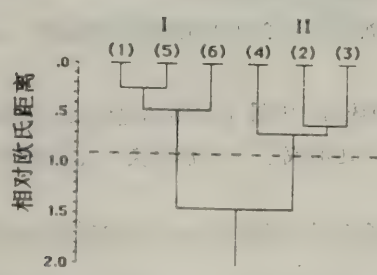


图 16.3 利用 RED 和可伸缩对策对六个巴拿马地点进行聚类分析的树状图。水平的虚线表示的是划定聚类 I 和 II 的人为划分线。

六个地点(SUs)的聚类方式在图 16.3 中的树状图中进行了总结。为了说明如何解释这些结果,我们人为地用了距离为 0.9 这个截断水平(象图 16.3 中显示的水平虚线)。在这个相似水平上,存在明显的两组: I (SU_s1,5,6)和 II (SU_s2,3,4)。根据表 11. 4a 可以看出, SU_s1,5, 6 很大程度上是由单种 *Latindia dohrniana* 占优势。当然,这是一个简化的数据集,仅仅用于说明之用,进一步解释 CA 结果可见 16.6 部分。

16.5 例子:威斯康星森林

表 11. 6a 中的威斯康星森林群落数据用 BASIC 程序 CLUSTER.BAS 进行了处理。利用

四种对策和弦距离(CRD, 14.9 部分)处理高地森林 10 个地点(8 种树)的 CA 结果, 在表 16.4 中给出。

可伸缩对策的结果在一个树状图中(图 16.4)作了总结。弦距离为 1.0 和 1.5 的两个人为截断水平(水平虚线)被用作区分聚类的参考点。在 1.0 这个距离可以分出三个聚类: I (SU_s, 1, 2, 3 和 4), II (SU_s5 和 7), III (SU_s6, 8, 9 和 10)。在更高的, 弦距离为 1.5 水平处, 聚类 I 和 III 融合, 形成了一个聚类。因此以刺栗和黑栎占优势的地点(SU_s1-4)形成了一个聚类和其余六个地点(SU_s5-10)相区别, 后者以美国椴和糖槭为特征(见表 11.6a)。

从每一个 CA 对策(表 16.4)的比较中可以看出, 矩心方法和组平均方法的结果基本上和可伸缩对策的结果一致。主要的不同是在 SU5 和 SU7 的聚类中。这两个 SU_s 实际上是介于聚类 I 和 III 之间(图 16.4), 即它们具有两个聚类的种类特征(见表 11.6a)

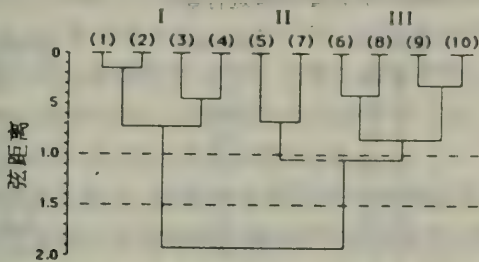


图 16.4 利用 CRD 和可伸缩对策对威斯康星南部高地森林 10 个地点进行聚类分析的树状图。
水平虚线代表限定聚类 I、II、III 的参考点

这些结果说明了很重要的一点。我们的经验表明, 当 D 矩阵中存在较明显的组时, 这些 CA 对策一般可以得到很相似的结果。威斯康星森林数据集合就很好地说明了这一点; 因为表 11.6a 中数据的简单性, 格局很明显, CA 的结果和这些观察一致。当数据集大而复杂, 且对于生态学家看来格局不明显时, 问题便会出现了。不同的对策可能得出不同的结果, 因为非常明显的聚类格局不可能必然出现于每一个对策中, 因此必须十分小心。注意这些 CA 结果和 AA(15.4 部分)的取样单位分类稍有不同, AA 中 SU_s1-5 形成一组和 SU_s6-10 分离(见图 15.2)。

表 16.4 程序 CLUSTER.BAS 给出结果(a)威斯康星森林 10 个地点间的弦距离和聚类, 利用的是(b)加权矩心, (c)不加权矩心, (d)组平均和(e)可伸缩对策

(a)弦距离 (D 矩阵)									
SU _s	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	0.15	0.61	0.50	0.83	1.17	1.02	1.22	1.35	1.30
(2)		0.64	0.50	0.85	1.16	1.02	1.22	1.33	1.29
(3)			0.45	0.94	1.14	1.12	1.18	1.38	1.25
(4)				0.57	0.95	0.90	1.05	1.28	1.18
(5)					0.79	0.67	0.95	1.16	1.07
(6)						0.74	0.41	0.80	0.80
(7)							0.63	0.62	0.61
(8)								0.52	0.53
(9)									0.31

(续前表)

(b)用矩心(加权)对策进行聚类

聚类轮数	组序	聚类水平	参考 SU ^a	组中的 SUs
1	9	0.15	1	2
2	8	0.31	9	10
3	7	0.41	6	8
4	6	0.45	3	4
5	5	0.41	1	2,3,4
6	4	0.48	6	8,9,10
7	3	0.44	6	7,8,9,10
8	2	0.62	1	2,3,4,5
9	1	0.62	1	来自一组所有 SUs

(c)用矩心(未加权)对策进行聚类

聚类轮数	组序	聚类水平	参考 SU ^a	组中的 SUs
1	9	0.15	1	2
2	8	0.31	9	10
3	7	0.41	6	8
4	6	0.45	3	4
5	5	0.41	1	2,3,4
6	4	0.48	6	8,9,10
7	3	0.44	6	7,8,9,10
8	2	0.62	1	2,3,4,5
9	1	0.65	1	来自一组所有 SUs

(d)用组平均对策进行聚类

聚类轮数	组序	聚类水平	参考 SU ^a	组中的 SUs
1	9	0.15	1	2
2	8	0.31	9	10
3	7	0.41	6	8
4	6	0.45	3	4
5	5	0.56	1	2,3,4
6	4	0.61	7	9,10
7	3	0.67	6	7,8,9,10
8	2	0.79	1	2,3,4,5
9	1	1.13	1	来自一组所有 SUs

(e)用可伸缩对策进行聚类 $\beta = -0.25$

聚类轮数	组序	聚类水平	参考 SU ^a	组中的 SUs
1	9	0.15	1	2
2	8	0.31	9	10
3	7	0.41	6	8
4	6	0.45	3	4
5	5	0.67	5	7
6	4	0.72	1	2,3,4
7	3	0.84	6	8,9,10
8	2	1.04	5	6,7,8,9,10
9	1	1.93	1	来自一组所有 SUs

a 组中最小数值的取样单位

16. 6 聚类分析的附加讨论

在本章中,我们介绍了四种利用 Lance 和 Williams (1967) 结合线性模式的等级聚类方法。一旦计算出距离矩阵(D),这些方法的计算效率都比较高,因为 D 矩阵中包含了全部用于 SUs 聚类的信息。而其它对策需要在每一轮内反复使用初始 D 矩阵,因此十分讨厌,并且计算效率较低。Anderberg (1973), Ganch (1982), Goodall (1978a), Orloci (1979), Pielou (1977, 1984), Romesburg (1984), Sneath 和 Sokal (1973) 和 Whittaker (1978b), 写了各种 CA 方法的大量评论文章。

另一种等级聚类方法是流行于生态学家之中的最小方差聚类(也称为 Ward 方法—见 Everitt 1974, Hartigan 1975, Orloci 1976a)。这种方法有很大的直观性,因为它是基于以下简单的原理:在聚类的每一步中,考虑聚类间方差的同时,使聚类内方差最小。组内方差被定义为聚类内各 SUs 和该聚类的矩心距离的平方和。在每一轮聚类中,方差最小的两个 SU 进行聚类相互结合后,方差增大(相对于每一个聚类内彼此独立的方差)。这种方法需要的计算可由用于 Ward 方法聚类过程的 SAS(统计分析系统)完成。

我们在 16.2 部分中简单介绍过,由聚类处理提供信息划分的同质群落,一般带有主观性。生态学家一般对给定数据集中可能包括的群体(取样单位组)数有所直觉,在可得到这个群体数的聚类水平上截取树状图(图 16.4)的茎是一件简单的事情。

然而,还有许多可以和生态学家的直觉结合使用的客观方法。最早提出的用于制作树状图的一种方法是共表型性相关(Sneath 和 Sokal 1973),此处由树状图暗示的取样单位间距离和原始 SU \times SU 距离阵相比较。进行每一轮聚类处理,形成越来越大的组(所有取样单位都结合在一起时为止),初始 D 矩阵距离和树状图距离之间的相关逐渐减小。从一轮到下一轮的相关,如果有一大的下降则表明应该在前一轮停止合并。例如,如果 CA 的共表型性相关(图 16.4)在弦距离是 1.0(上虚线)时,为 0.80,当弦距离为 1.5(下虚线)时为 0.5,那么也许在弦距离为 1.0 截断得到的聚类便是同质群体。

Orloci (1976a) 提出,在最小方差聚类过程中,可以决定划分同质组的显著度水平。Goodall (1978a) 提出了一个大概的方差比率,用它来检测由于 SU 合并或 SU 组合并引起的方差增大是否处于一个“可接受的”(例如 $P=0.05$) 水平内。Ratcliff 和 Pieper (1981) 对这种方差分析方法作了概括,它包括一个假设的检验,该假设认为平均组内距离和平均组间距离没有显著的差别。他们的过程开始于在两组水平上应用检验,即检验把所有 SUs 当成一组和分成两组时,各 SU 之间平均距离的差异。该方法和 Hill(1980)的停止规则过程相似。关于估测分类最近的其它进展,我们推荐同学们看一下 Archie (1984) 和 Rohlf (1974) 的文章和 Duncann Estabrook (1976), Popma 等(1983), Rohlf (1982) 的研究。

表 16.1 中显示的 CA 对策(矩心、组平均和可伸缩)被认为是空间保持的;在各种(距离)水平上的 SUs 聚类,当这些聚类距离和初始 SU \times SU D 矩阵距离比较时,引入的畸变很小。

我们介绍的并未包括 Lance 和 Williams 线性结合方程中使用的所有对策。例如,简单连锁(Single linkage)和完全连锁(complete linkage)对策(Sneath 和 Sokal 1973)就被省略了。不论是用简单连锁缩小距离,还是用完全连锁扩大距离(Pielou 1977),这两种 CA 方法的空间畸变

都很强烈。即合并后重建的距离矩阵和初始距离矩阵(D)有很大的差异。

矩心对策经常造成结果的颠倒；一些实体对之间的矩心距离可能会小于上一轮内某一对之间的矩心距离。如果参数 α_1, α_2 和 β 之和等于 1，那么以后的等级结合将是单调的，且不会发生颠倒。因此，因为参数之和限定为 1，可伸缩对策可定义为单调的。可伸缩对策的主要特点是， β 控制着聚类对策的空间保持性，变化 β 则空间可以扩大或缩小。当 β 接近 -0.25 时，倾向于空间保持，但当 β 负值更大（即其绝对值增大）时，畸变将扩大；而 β 正值更大时，畸变缩小。我们建议同学们查阅 Sneath 和 Sokal (1973) 的文献，以便进一步了解可伸缩对策的细节。

在这一章中我们介绍了聚类分析(CA)，假设这些数据来自于景观中随机散布的 SUs，数据是种的丰富度或种的有无。而且，若聚类的取样单位数据来自一个时间序列，则可用于检测生态演替模式(Legendre 等 1985)。其它类型的观察数据也可用于聚类，例如，森林树木大小的等级或土壤剖面数据(Faith 等 1985)。

16.7 总结和推荐

1. 聚类分析(CA)是基于实体(SU)相互之间总的相似性，把实体分成组或聚类的一种技术。相似的 SU 将形成聚类与其它 SU 聚类分离。聚类分析是一个一般性的术语，它涉及到许多这样的规则系统，这些规则系统间的主要区别是他们对聚类结构的处理。

2. Lance 和 Williams (1967) 的 CA 总规则系统是一个线性结合方程[方程(16.1)]。方程 16.1(如表 16.1 中显示)的参数选择不同值，可以形成四种对策：矩心(加权的和不加权的)，组平均和可伸缩的。

3. CA 的结果可以很方便地概括于一个树状图中(例如，图 16.2)。从这个树状图中识别出明确的组或群体，多少带有些主观性。作为总的指导方针，我们提倡不要分的如此之细，以至使结果产生大量过于破碎的聚类。一些研究者曾提出过一些客观的方法(16.6 部分)。虽然利用 CA 这样的方法帮助解释数据，对同学们是有益的，但如果过分强调这个简单的分析结果就错了。

4. 当被分析的基础数据集具有一些相对明显的格局(例如，16.5 部分中的示例)时，大多数 CA 对策得到的结果是相似的。但当数据集大而复杂；且无明显格局时，各种 CA 对策的结果常常不相同(实际是某些情况下)。在后一种情况下，我们推荐选择几个对策，并比较它们的结果，这样的比较常对识别合乎逻辑的聚类有帮助(考虑到数据潜在的生态学意义)。

第六部分 群落排序

第十七章 背景

近年来统计生态学得到重大发展的一个领域就是群落排序,排序是一个术语,它是用来描述一组技术,在这组技术中取样单位(SUs)被安置在与一个或多个相对应的坐标轴上,这样它们相对于轴或其它的位置可以提供关于它们生态学上的相似性的最大信息。依据SUs的相对位置,通过识别它们之间的相似(相异)程度,我们也许可以找到引起观测到的格局产生的一些潜在因子。最终目的就是通过推论去阐明那些生物的和环境的因子在测定样本所来自的生态群落结构时可能是重要的。排序的目标就是简化和压缩众多的数据以期待一定的生态关系出现。Orlaci(1978)论述了排序的目的就是“概括”。

从概念上讲,可以把排序设想为在一个种的多维空间内放置取样单位,也就是在一多维空间,每个种都有单一的维数或坐标轴(见图 13.2),最简单的一个例子,假定一单一的成分比如说一个种的多度A,在一个样本中的每个SU中被测定。现在这些SUs就可以沿着这条单一的多度轴排序(即种的多度从0到100)。位于排序轴相反两端的SUs当然是最相异的,那些在这条轴上紧接近的SUs是最相似的,这个图示方法可以扩展到同时有两三个种的情形。很明显,特别是当有许多种和SUs的时候这就很快变得难以处理。

排序的目标应当同分类的目标明显地区别开来(15章,16章),对于一给定S行(种)和N列(SUs)的矩阵X,分类的目标就是减少 $X_{s,n}$ 为 $g < N$ “同类的”群丛组,也就是g个群丛组的每个群丛内的SUs之间的相似性要高于不同群丛间的SUs之间的相似性(Green 1983),而另一方面,排序的目的是完成转化 $S \times N$ 矩阵并减少它为 $P < S \times N$ 矩阵,这个转化通常是指“减少维数”(Legendre 和 Legendre 1983),很明显,如果我们有很多种而且可以减少维数到几个(比如说 $P=2$),那么SUs之间的关系就比较容易测定,排序的目的之一就是在导出这些P维的同时使信息的损失减少到最低程度。

Goodall(1954b)介绍排序这个术语为去认识与“多维空间系列”有关的用于排列植被样本(或种)的一系列技术(Austin 1985),Goodall的工作在植被作为一连续体概念数量化的需要下得到发展,也就是植被样本或取样单位表示的是来自真正“连续的”世界的一个样本(连续体的和个体的概念——Glenson 1926, McIntosh 1965)。这两个互相补充的概念与构成分类基础的生物群落的间断体和有机体的概念相对照(尽管一些观点认为这对照是“不存在”的问题——见Anderson 1965)。尽管由排序而来的一些抽象的格局,可能表示的是间断的群落(即在坐标系内的SUs组)。但通常排序的产物并不是分类的产物,一个看上去很自然的,有用的群落分类可能“从属于”一个排序的结果(见Jenson 1979),而且排序可能应用到由分类划分的大数据组的一个亚组。

排序的两个方法可以区分为:(1)直接的,(2)间接的(Gauch 1982, Whittaker 1965)。在直接排序中,SUs沿着作为研究基础而事先选择测定的环境梯度放置(Austin 等 1984),尽管本

书没有包括直接排序的方法,但在第 20 章给出了这一方法总的展示,而间接排序,SUs 根据它们在种类组成上的相异性(或相似性)排列在减少的坐标系统上,这样 SUs 间通过合成轴或梯度建立关系,因而这种排序被视为间接的(Whittaker 1967,1978a)。从第 18 章到第 20 章里我们将描述几个间接排序的方法。

历史上,除了 Goodall(1954b)的工作外,生态群落排序技术的早期发展可以追溯到 Wisconsin 植物生态学家 John Curtis 和他的学生们(如 Brown 和 Curtis 1952,Curtis 和 McIntosh 1951)早期的工作。由于早期的一些单一轴排序不能令人满意,因而导致了由 Bray 和 Curtis 在 1957 年发表的一种多轴方法的发展,这种方法众所周知称为 Bray-Curtis 或威斯康星排序(Wisconsin ordination);现在一般称为极点排序(Polar ordination,PO)(Beals 1984,Gauch 1982,Orloci 1978)。极点排序的过程将在第 18 章里描述。

另一个多轴排序方法,由 Goodall(1954b)最先在生态学中使用,是主分量分析(Principal component analysis,PCA)。最初 Goodall 关于 PCA(他称为因子分析,Factor analysis)的文章很少有人注意,但随着计算机的发展,PCA 成了在排序中广泛使用的一种排序方法。将在第 19 章里描述 PCA,但 PCA 有些局限性,特别是当在处理具有很强的非线性数据时。对应分析(Correspondence analysis)技术在某种程度同 PCA 相比处理非线性数据具有较强的能力,将在第 20 章里描述。最后,两个非线性方法,多项式排序(Polynomial ordination)和非度量多维调节(Nonmetric multidimensional Scaling)在 21 章讨论。

Austin(1985),Beals(1981),Gauch(1982)和 Whittaker(1978a)提供了关于排序的概念和目的的精彩讨论。我们提供这些参考文献供那些除了对本书以外的一些排序方面东西感兴趣的同学参考。

Whittaker 和 Gauch(1973)列出了排序过程中的下列的一些成分。它们也许可以作为评价排序的标准:(1)来自畸变的相对自由,(2)结果的清晰或明了,(3)数据的有效利用,(4)对于揭示其它没有被认识到的格局和关系有启发价值。

17.1 矩阵展示

排序的过程以检测某个群落数据矩阵中种和 SUs 总的关系而开始(图 17.1)。

		取样单位									
种	a										
	b										
	c										
	.										
	e										
因子	w										
	x										
	.										
	z										

图 17.1 阴影部分表示的是用于排序的生态数据形式,SU 排序可以以种间的 R 状况相似性(即行之间)或 SUs 间 Q-方式的相似性为基础

根据使用的特定方法,计算成对SU_s间或成对种间的相似性。由这些SU或种对的比较矩阵可以导出排序的坐标系统。回忆一下排序的目的就是导出表示取样单位间关系的坐标轴系统,如果SU_s排序以种间R-方式的相似性为基础(即行),它指的就是R方案,如果取样单位排序的计算利用的是Q-方式的相似性则指的是Q-方案(Noy-Meir 和 Whittaker 1978, Orloci 1978)。

17.2 选择的文献

生态学家使用的这些排序方法表现了由早期极点排序的使用到现在强调的主分量分析,无偏对应分析(Detrended correspondence analysis)和非度量多维调节(Nonmetric multidimensional scaling)(表 17. 1)的自然发展过程。当生态数据矩阵在种间具有非线性关系时,我们推荐使用:(1)试图对非线性进行校正的排序方法(如:无偏对应分析——Hill 和 Gauch 1980),(2)围绕线性假定的排序方法(如:非度量多维调节——Kenkel 和 Orloci 1986)。

表 17. 1 使用极点排序(PO)、主分量分析(PCA)、对应分析/无偏对应分析(DCA)、或非度量多维调节的生态排序研究的作为选择的文献一些例子

位置	群落	方法	参考
威斯康星	高地森林	PO	Bray 和 Curtis 1957
威斯康星	高草草原	PO	Dix 和 Butler 1960
威斯康星	林鸟	PO	Beals 1960
渥太华	海底生物	PO	Erman 1973
英格兰	白垩草地	PCA	Austin 1968
安大略湖,加拿大	云杉林	PCA	Jeglum 等 1971
萨斯克,加拿大	森林地衣	PCA	Jesberger 和 Sheard 1973
加纳	蚂蚁和植物	PCA	Majer 1976
得克萨斯	火蚁	PCA	Pimm 和 Bartell 1980
北大西洋	浮游植物	PCA	Matta 和 Marshall 1984
澳大利亚	山地草地	PCA	Foran 等 1986
Hoy Orkney	草地灌丛	MDS	Prentice 1977
挪威	高寒草甸	MDS	Matthews 1978
亚尔贝特,加拿大	大型无脊椎动物	DCA	Culp 和 Davies 1980
英格兰	沼泽	DCA	Hill 和 Gauch 1980
芬兰	北欧石楠	DCA,MDS	Oksanen 1983
葡萄牙东南	半干旱	DCA,MDS	Dargie
威尔士	沙丘草地	DCA	Gibson 和 Greig-Smith 1986

第十八章 极点排序

极点排序(PO)是最早广泛地被植物生态学家应用的方法之一,尽管由于 PO 有些局限性,同时较多的数学的高级排序方法正在被利用(第 19、21 章),PO 还是在不断地产生一些具有意义的生态结果,事实上,PO 经常作为比较新方法的一个标准,在这章里,我们描述 Bray 和 Curtis(1957)的极点排序方法。

18.1 基本方法

Bray 和 Curtis 的 PO 方法是被特定发展以分析植物群落数据的较少技术之一(Beals 1984),PO 的目标是将 SU 放在一个坐标系统内,以取样单位(SUs)间的距离反映基本的相似性。同时也希望它们与潜在环境梯度的关系得到反映,基本的步骤包括选择 SUs 作为在一个坐标轴上的端点(极点),接着是对剩下的 SUs 相对这些端点的一个简单的几何放置。生态学家对于端点 SUs 的环境不同性经常有一种较强的直觉,因而此轴就代表了梯度。

PO 最显著的特征就是简单性。它的计算步骤直截了当,所以同较高级的数学的排序过程(如主分量)相比,具有很少的威胁性。

18.2 步骤

下面的步骤基本是来自 Cottam 等(1978)对原初 Bray 和 Curtis 方法的描述。

Bray—Curtis 极点排序

假定有 N 个 SUs,在每个 SU 感兴趣的是对 S 种每个多度(如密度、百分盖度、生物量)的一些测定。

第一步. 计算 SUs 间的相似性。正如第 16 章描述的分类步骤情形那样,PO 的第一步是计算 SUs 对 Q —方式的相似性,Bray 和 Curtis 用百分相似性[PS 公式(14.6a)]作为它们相似性的测量[PS 有时也称 Czekanowski 指数(Bloom 1981, Goodall 1978b)],回忆一下第 14 章,PS 的余数相异性(即 $100 - PS$)称为百分相异性(Percent dissimilarity)[PD 公式(14.6b)],尽管 Orloci(1978)不推荐在欧氏类型(Euclidean-type)排序中使用 PD(即那些需要严格的几何步骤),但在这我们还得使用它,因为传统的 PO 用的就是它。在 PO 中使用 PD 既有很强的拥护者(Beals 1984),也有反对者(Orloci 1974b, 1978),这一问题我们将在第 18.6 节进一步讨论。

按第 14 章(公式 14.6b)的阐述计算每对 SUs 间的百分相异性。

第二步. 构造 $SU \times SU$ 相似性矩阵。所有对 SUs 的 PS 和 PD 值被总结在这个 $SU \times SU$ 矩阵形式中,对角线上方为 PS 值,对角线下方为 PD 值。

第三步. 在 X 轴上选取端点 AX 和 BX 。对每个 SU 计算它与其它每个 SU 相异性总和。

(a)相异性总和最高的定为 AX 。

(b)相对于AX具有最大PD值的SU定为BX(选择BX时,如果同AX的百分相异性(PD)大于95%的不只一个SU,那么选择相异性总和最大的SU)。

(c)X轴的长度(L)由AX和BX间的PD值给出(即 $PD_{AX,BX}$)。

第四步.在X轴上放置剩余的SU。相对于端点的SUs AX和BX的剩余SUs在X轴上的位置(x)由Beals(1965)公式计算。

$$x(i) = \frac{L^2 + dA(i)^2 - dB(i)^2}{2L} \quad (18.1)$$

$x(i)$ 是第*i*个SU在x轴上的位置,L是AX和BX间的PD值。 $dA(i)$ 是第*i*个SU对于AX的PD, $dB(i)$ 是第*i*个SU对于BX的PD,公式18.1图示在图18.1上。

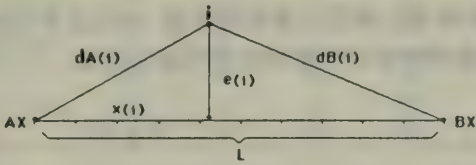


图18.1 在极点排序中在X轴上(由端点SUs AX和BX确定的)第*i*个SU放置的Beals公式的几何表示。符号的定义见公式(18.1)和(18.2)

第五步.选择Y轴两个端点AY和BY。在图18.1注意点*i*到X轴距离由勾股定理给出。

$$e(i) = \sqrt{dA(i)^2 - x(i)^2} \quad (18.2)$$

由图18.1可以看出,最大的 $e(i)$ 值出现在与两个端点SUs相异性等分的SU,这样就导致这个SU与AX和BX有等同距离的趋势(中轴处),为了将剩余取样单位放在Y轴上,就要考虑剩余这些SUs的偏离的最大成份,因而就需计算每个SU的 $e(i)$ 。

(a)具有最大 $e(i)$ 值的SU定为AY,选择AY的一个附加的条件就是它必须落在X轴中间范围的50%以内,也就是0.25L和0.75L之间。

(b)与AY具有最大PD值的SU定义BY,选择BY的进一步条件是在X轴上BY必须在AY的10%范围内,这保证Y轴将基本垂直于x轴[如果不止一个SU对于AY的PD值大于等于95%(其它方面附合上述条件)具有最大总和PD值的SU选为BY]。

(c)Y轴的长度为AY和BY之间的PD值(即 $PD_{AY,BY}$)。

第六步.在轴上放置剩余的SUs 相对于端点SUs AY和BY,剩余SUs在Y轴上的位置(y)通过恰当地代入Y轴的端点和长度利用公式(18.1)计算。[第三个轴(Z),也可以按着与X,Y轴描述的相同步骤得到计算]。

第七步:PO的图形显示。利用计算的 $x(i)$ 和 $y(i)$ 值,将每个SU放置在这个X-Y轴系统内,产生一个二维的PO。

18.2.2 简化的极点排序

PO早期应用在相对相似的生物群落数据中,如威斯康星高地森林(Bray和Curtis 1957),但生态学家经常用它处理来自许多不同类型环境的数据组,如Swan和Dix(1966)对高地和低

地森林的研究。当 PO 被用在这些“异类的”数据时,在选择 SU_s 端点时会出现一系列的问题,众多的 SU_s 对间可能是 100% 的不相似,而且很有可能不同对的 SU_s 将与不同的潜在环境梯度有关。这样就有哪一对 SU 应被选出作为第一排序轴的端点的问题。

在许多情况下,生态学家有关于存在的环境梯度有道理的先知,用这些先知可以解释不同群落种差异的格局,Cottam 等(1978)和 Gauch(1982)的讨论认为在 PO 的排序中,每个轴端点的选择应该是人为的,通过选择这些代表环境梯度的“极点”,这样以两个不同的 SU 在一起作为端点的问题就会减少,此外还有计算上的优点,即整个 SU 相异性的比较矩阵就不需要了,只需要被选出的两个端点 SU_s 同其它 SU_s 间的相异性。在 18.3 节的计算例子中,传统的 PO 计算如上面 18.2.1 节中所描述的将被阐述。但在 18.4 节计算机计算的那个例子中,简化的 PO 步骤将被使用,结合简化的 PO,利用计算机的优势,运行各种“不断摸索”的选择端点对排序结果的迅速检验,以选出最清晰和具有最高启发价值的结果。

18.3 举例:计算

下面计算阐述的 PO 步骤,使用的是在 5 个 SU 中 3 个种多度的数据矩阵和每个 SU 多度值的总和(表 18.1),尽管在许多研究中,使用 PO 时,在开始之前对数据实行“双标准化”,象 Beals(1984)和 Cottam 等(1978)讨论的那样,我们阐述 PO 的计算过程使用没有转化的数据。双标准化包括:首先以与最大多度的相对值表达每个种的多度(即行标准化),接着用这些转化了的值除以每个 SU 这种值的总和(即列标准化),这有许多效果,但主要是避免在 PO 的计算中几个少数高丰盛度种强的加权(Beals 1984)。

表 18.1 五个 SU_s 中种(Spp)多度,平均数和每个 SU 多度总和

SPP	SU _s					平均数
	(1)	(2)	(3)	(4)	(5)	
(1)	2	5	5	3	0	3.0
(2)	0	3	4	2	1	2.0
(3)	2	0	1	0	2	1.0
总数	4	8	10	5	3	

步骤 1. 计算 SU_s 之间的相似性。 第(i,k)对 SU_s 间的百分相似性[PS,公式(14. 6a)]和百分相异性[PD,公式(14. 6b)]如下:

$$\begin{aligned} PS(1,2) &= [2 \times (2 + 0 + 0) / (4 + 8)] \times 100\% = 33\% \\ PS(1,3) &= [2 \times (2 + 0 + 1) / (4 + 10)] \times 100\% = 43\% \\ &\dots \dots \dots \\ PS(4,5) &= [2 \times (0 + 1 + 0) / (5 + 3)] \times 100\% = 25\% \end{aligned}$$

和

$$\begin{aligned} PD(1,2) &= 100\% - 33\% = 67\% \\ PD(1,3) &= 100\% - 46\% = 57\% \end{aligned}$$

$$PD(4,5) = 100\% - 25\% = 75\%$$

步骤 2. 构造 SU×SU 相似性矩阵。在表 18.2 中给出了百分相似性相异性矩阵以及每个 SU 的 PD 总和。

表 18.2 五个 SUs 间 PS(对角线上)和 PD(对角线下)值的矩阵,和每个 SU 的 PD 总和

		PS—百分相似性					
		SUs	(1)	(2)	(3)	(4)	(5)
PD—百分相异性	(1)	—	33	43	44	57	
	(2)	67	—	89	77	18	
	(3)	57	11	—	67	31	
	(4)	56	23	33	—	25	
	(5)	43	82	69	75	—	
PD 总和		223	183	170	187	269	

步骤 3. 在 X 轴上选择端点 AX 和 BX。SU5 具有最大的相异性总和(PD=269),故被定为端点 AX, SU2 与 SU5 具有最大的相异性(PD=82%),故选作为端点 BX, X 轴的长度(L)为 82。

步骤 4. 在 X 轴上放置剩余的 SUs。用公式(18.1)和 L=82, SU1 在 X 轴上的位置是

$$\begin{aligned} x(1) &= [(82)^2 + (43)^2 - (67)^2] / (2 \times 82) \\ &= (6724 + 1849 - 4489) / 164 = 25 \end{aligned}$$

类似地,对于 SUs 3 和 4 有

$$\begin{aligned} x(3) &= 69 \\ x(4) &= 72 \end{aligned}$$

步骤 5. 选择 Y 轴端点 AY 和 BY。首先用公式 18.2 计算剩余 SUs e(i)值。

$$e(1) = \sqrt{[(43)^2 - (25)^2]} = \sqrt{1224} = 35$$

类似地,对于 SUs 3 和 4

$$\begin{aligned} e(3) &= 0 \\ e(4) &= 21 \end{aligned}$$

SU1 具有最大的 e(i)值,且落在 X 轴的 25 处(位于 X 轴中部 50%的范围内),故选作为 AY 端点,但没有其它的取样单位在 AY 的 10%范围内,因此本例中的排序被限定在单一轴上。

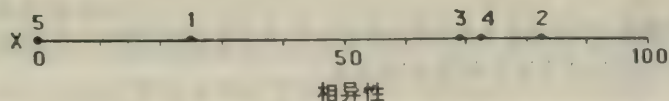


图 18.2 五个 SUs 以相异性百分比为基础,单一轴 Bray—Curtis 极点排序

步骤 6 PO 的图形表示。单一轴 PO 的结果表示在图 18.2 中,注意沿着 X 轴 SUs 的位置与对这些同样的 SUs 进行的聚类分析的顺序是十分相似的(图 16.2)。

18.4 例子:巴拿马蟑螂

简化的 PO 步骤(18.2.2 节)被应用到 5 种蟑螂在巴拿马 6 个发生地的多度数据上(表 11.4a)利用程序 PO.BAS(软盘内)进行排序,尽管这个 PO.BAS 还提供了其它距离方法,但我们还是结合 PD 作为相似性测量。

表 18.3 使用 BASIC 程序 PO.BAS 对巴拿马六个发生地蟑螂极点排序结果的总结

输入 X 轴端点 SU 数码? 4,6		
输入 Y 轴端点 SU 数码? 1,5		
端点 SU 4)和其它 SU 百分相异性		
1)=98.3 2)=97.3 3)=33.3 5)=88.2 6)=100.0		
端点 SU 6)和其它 SU 百分相异性		
1)=66.2 2)=50.0 3)=100.0 4)=100.0 5)=46.3		
端点 SU 1)和其它 SU 百分相异性		
2)=50.0 3)=96.6 4)=98.3 5)=75.9 6)=66.2		
端点 SU 5)和其它 SU 百分相异性		
1)=75.9 2)=63.2 3)=77.8 4)=88.2 6)=46.2		
二维极点排序 X 和 Y 坐标		
SU	X	Y
(1)	76.4	0.0
(2)	84.8	28.1
(3)	5.6	59.6
(4)	0.0	50.3
(5)	78.2	75.9
(6)	100.0	52.6

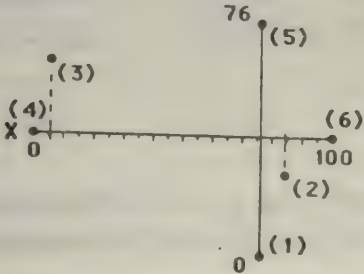


图 18.3 5 种蟑螂在巴拿马 6 个发生地的极点排序,按着它们端点的 SUs 划出 X 和 Y 轴,以强调 SUs 作为每个轴“极点”的概念

注意,由表 18.3,PO. BAS 鼓励使用者选择 X 和 Y 轴端点。由表 11.4a 我们可以发现 SU6 与 SU3 和 SU4 没有共同的种,因此有理由——将这些 SUs 组合(即 SUs6-3 和 6-4)作为 X 轴端点试一试,SUs6-4 的选择给了我们 X 轴上 SU 的明确分离(图 18.3)。对于 Y 轴,我们试了剩余的几组组合,并得出 SUs1 和 5 提供了最清晰的结果(对于这个特殊的例子,这种“不断摸索”的方法,给了我们与用完整的 Bray-Curtis 的 PO 过程相同的结果)。

在图 18.3 中,6 个 SUs 被排列在两个坐标轴系统的空间上,它反映了它们完整的相似性,由于原来的数据矩阵具有 5 个种和 6 个 SUs,而现在转化为一个 2×6 矩阵,所以我们获得了减少的维数。尽管这个排序反映了所有种的多度,但优势种(*Latindia dohrniana*)对于 PO 有很大的影响,在 X 轴端点的一侧 SUs 具有较高的多度(SUs1,2,5 和 6),在另一端 SUs3 和 4 的多度是 0(图 18.3)。

当然,在最终的分析中,任何 PO 的价值在于它们的启发性。它能否提供一个有用的生态解释?X,Y 轴反映没反映种多度格局和潜在环境因子。这些重要的问题和提出他们的方法,我们将在第 24 章描述。

表 18.4 极点排序结果:(a)PD 矩阵和总和,(b)10 个森林立地(SUs)每一个的 X 坐标和离差,使用的是 PO. BAS

(a)百分相异性矩阵和总和										
SUs	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	—	7.4	29.4	25.4	43.9	70.4	56.9	71.9	81.5	75.4
(2)	7.4	—	33.3	25.4	43.9	70.4	53.8	71.9	77.8	75.4
(3)	29.4	33.3	—	17.9	48.1	72.5	64.5	77.8	92.2	77.8
(4)	25.4	25.4	17.9	—	25.8	55.9	42.9	61.3	79.7	67.7
(5)	43.9	43.9	48.1	25.8	—	44.4	32.4	53.5	71.9	60.0
(6)	70.4	70.4	72.4	55.9	40.4	—	35.4	19.3	40.7	43.9
(7)	56.9	53.8	64.5	42.9	32.4	35.4	—	29.4	32.3	26.5
(8)	71.9	71.9	77.8	61.3	53.5	19.3	29.4	—	22.8	23.3
(9)	81.5	77.8	92.2	79.7	71.9	40.9	32.3	22.8	—	15.8
(10)	75.4	75.4	77.8	67.7	60.0	43.9	26.5	23.3	15.8	—
总和	462.2	459.4	513.5	402.0	419.6	448.9	374.1	431.1	514.6	465.8

(b)在 X 轴上 SUs 坐标和离差										
SUs:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
x(i):	77.4	72.9	92.2	78.8	61.6	26.5	29.2	16.1	0.0	14.6
e(i):	25.4	27.2	0.0	11.8	37.2	30.9	13.9	16.2	0.0	6.0

18.5 例子:威斯康星森林

利用 8 个树种在南威斯康星森林立地的多度数据(表 11.6a),再一次利用 PO. BAS 计算 PO(表 18.4)。SU9 与其它 SUs 的 PD 值总和最大(574.6),故选作为 AX,SU3 与 SU9 具有最大的 PD(92.2),故选为 BX。

确定了剩余 SUs 在 X 轴上的坐标后,计算每个 SU 到 X 轴的离差值 e(i)(表 18.4b),SU5 与 X 轴的离差最大(37.2)且落在 X 轴的中间范围内。因此,选为 Y 轴的第一个端点(AY)。但没有其它的 SUs 在 SU5 的 10%范围内,由于作为 Y 轴端点的标准得不到满足,因而就不用再

计算 Y 轴了(注意 SU2 在 SU5 的 12% 的范围内,但如用它作为端点 BY,在 X 轴和 Y 轴之间将产生一个不可接受的斜角,由于 X 轴和 Y 轴代表独立的环境梯度,X 和 Y 轴之间应该是垂直的或近似于直角的)。

10 个森林立地沿 PO 的 X 轴排序图表明 SUs1,2 和 4 的位置与端点取样单位很近,表明它们与 SU3 的相似性很接近(图 18.4),在这些 SUs 中,栎树和黑栎树为优势种,SUs6,7,8 和 10(糖槭和椴树占优势)与端点 SU9 很接近,因此它们与 SU9 很相似。

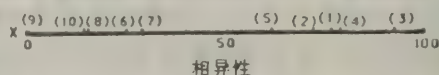


图 18.4 威斯康星高地 10 个森林立地单一轴(X)的极点排序

18.6 极点排序的附加讨论

Sneath 和 Sokal(1973)指出使用 Q-方式相似性方法的优点在于,它允许 SUs 之间的关系以几何的形式具体化,但这个几何空间需要的不是欧氏距离,它应该是可度量的(Metric),这样空间的拓扑结构才能被相似函数单独测定,一个令人满意的公制条件是三角不等:

$$D(j,k) < D(j,l) + D(k,l) \quad (18.3)$$

在这一章里我们用来测定 PO 的百分相异性(PD)在一定情形下对于这个条件不成立,如对表 18.2, $D(2,5)=82$, $D(2,3)=11$, $D(3,5)=69$, 而 $D(2,5) > D(2,3) + D(3,5)$ 或 $82 > 11 + 69$, 它们对于公式 18.3 不成立,由于几何方法在 PO 中的重要性,Orloci(1974b)推荐全部用可度量相似测量,Deals(1984)指出,如果数据进行标准化,那么存在于 PO 的这个三角不等问题就消失了(见下面)。BASIC 程序 PO.BAS 提供了几个距离函数的选择,我们鼓励学生们用 PD 和其它的 1,2 个方法运行这个程序,并检验这些结果的“生态一致性”,我们建议查一查这种不同性是由方法、数据还是两者结合造成的。

Orloci(1974b, 1978)阐述了 PD 的非度量成分,并指出它的潜在引起原因是出在它的除数 PS 上,[公式(14.6a)]。它是一对样方所有观测值的总数,由于对于每对 SUs 这个总数可以不同,每个 PS 值可以建立在不同度量(除数)上,此而失掉了可度量成分,Beals(1984)注意到如果数据进行列标准化,即对于给定 SU 每个种的多度被这个取样单位多度总和除,那么除数将是恒定的,PS(PD 的补充成分)将是可度量的,Beals 进一步指出“在文献中的所有证据表明, Sorenson 系数(PS)同许多可度量方法相比,如欧氏距离在多元分析中给出了较多的生态上可以解释的结果”。

PS 的一些统计成分是可知的,Richlefs 和 Lau(1980)通过计算机对 PS 的分布(标准差)特征与不同样本容量的模拟,得到正如预料的结果那样,小容量的样本标准差很大,同时发现, SUs 间的相似性较高(接近 100%)或较低时(接近 0%)标准差趋向于最小。

正如 Wisconsin 森林例子显示的那样(18.5 节),极点排序存在的一个问题就是,因为 Y 轴的最佳端点 SUs 在 x 轴上经常偏离一段距离,所以 Y 轴对于 X 轴可能是斜的,这些排序轴不能完全代表独立的梯度,这种斜的问题对于大的数据组并不是常见的,因为对于大的数据组, Y 轴端点 SUs 具有较多的选择,Orloci(1974b, 1978)提供了校正不垂直的 PO 轴分析和图示的详细解决方法。

生态学家面临的另一个问题就是非常异类的数据,这类数据的一个固有特征就是种—多度关系强烈地趋向于非线性(这就是我们经常所说的非线性数据结构),所有的排序方法都受到这种非线性不同程度的影响(Kenkel 和 Orloci 1986)Cottam 等(1978)表示威斯康星的极点排序在处理非线性数据时“决不是十分安全的技术”。非线性是排序理论和方法的一个主要问题。我们推到第 21 章讲述这个问题。

这两个问题(端点 SU 选择和非线性),和其它少数几个严重问题导致了众多的对原始 Bray 和 Curtis PO 的修改(Bannister 1968,Beals 1984,Gauch 和 Scruggs 1979,Orloci 1974b),生态学家修改了这种基本的技术使之能够适应他们数据的特殊性。尽管这展开了生态学家对人为的(选择 SUs)的批评,但我们感觉生态学家们最了解自己的数据和怎样最好地从这些数据中提炼出具有意义的生态格局(在使用 PO 这种情况)。我们鼓励学生按着 Beals(1984)建议的方向修改 PO 选择端点 SUs 的标准,特别是当在处理非常异质的群落数据时。这些异质性的问题也是与选择端点 SUs 有关——注意在选择一个“分离物”作为端点时,其它的 SU 趋向于在“分离物”SU 的另一侧聚集。

18.7 总结和推荐

1. 极点排序有吸引力是因为它相对于较多数学的排序方法具有步骤和概念简单的特点。尽管经常受到对它的主观性的一些批评,我们还是与 Beals 的观点一致即 PO 是一种有价值的方法。

2. 在 PO 的每一步,常常需要一些相当主观的决定(如数据应不应转化;使用哪种相异性方法?哪一个 SU 作为端点 SUs),但,PO 如此成功地在生态学研究中使用主要原因之一就是生态学家可以根据这种方法的需要直截了当地作出过程决定。一些其它的排序方法趋向为一种“黑箱子”,使用者很大程度上与分析脱离。

3. PO 中重要的一步是 SUs 对间 Q—方式相似性方法的选择(18.2.1 节),对每组数据选择一种恰当的相似性方法有时是困难的。Orloci(1974b)推荐用可度量方法(如弦距离),它们给出的重要性 with SU 对间种的多度“相对”不同(18.6 节),Beals(1984)推荐用 PS。我们建议几个方法都试一下并比较其结果,如果在某组数据中有个正确的明显的格局,每种分析将一致地给出。

4. 在端点选择的重要一步中,生态学家必须建立选择标准,这个标准将能导致一个有生态意义的轴产生。尽管她/他可能由于主观性而受到批评,当选择端点以反映潜在格局梯度时,我们认为生态学家总是有可利用的最好的先知。

5. 我们建议学生避免选择“分离物”作为端点 SUs 因为它可能引起其它 SUs 在相反的端点处聚集。Gauch(1982)和 Beals(1984)提供了与端点 SU 有关的问题的进一步处理。

6. 对于非常异质性的数据,强的非线性种—多度关系可能严重地歪曲出现在 PO 中的 SUs 格局(或者任何的其它排序方法;Kenkel 和 Orloci 1986)。在这种情况下,我们建议将这些数据分成较多的同质亚组(可能通过排序方法),或生态学家应用非线性的排序方法(21 章),如果使用 PO,应当注意可能出现的畸变类型。

第十九章 主分量分析

主分量分析(PCA)有段时间是在生态学中最广泛使用的一种排序过程,它最初吸引人的地方是它表面上(或许它的一点透明性)数学化形式和在计算机统计软件包上的可利用性,生态学家可以简单地把他或她的数据“塞入”这些软件包之一,并让计算机计算SU排序坐标,不需要作出象哪种相似性方法是确切的,或者怎样选择端点SUs(象在极点排序中那样,第18章)的困难决定。随着时间的推移,作为一种排序技术,PCA的问题就变得明显了,并且人们正在寻找新的排序技术。但我们在它的限定范围内使用时,PCA不失为一种有效的排序方法,这里我们提供PCA方法,同时讨论一下它的优缺点。

19.1 基本方法

Goodall(1954b)最先在生态学中应中PCA,但它的发展可以追溯到Pearson(1901)的文章“空中航线和飞机对于点系统的最佳配合”。PCA基本上是一种用来处理矩阵内部结构的多元统计技术。我们的经验是学生们通过一些例子可以对PCA技术方面的一些东西得到很好的理解。下面给出PCA简单技术方面的一个总的轮廓,我们希望学生们通过以后几节(19.2—19.5)的工作(学习)再重新阅读这一轮廓。通过一个例子的理解后,PCA难懂性的威胁就会无疑地减少。

PCA是一种打破或分离相似性矩阵为一组相互垂直(或直角的)轴或分量,传统上,这个矩阵是由方差—协方差或相关组成,但PCA也可以应用到欧氏矩阵中(Gower 1966, Orloci 1973)。每个PCA轴对应于矩阵的一个特征值。PCA也经常是指作为特征分析排序方法之一(Orloci 1978, Pielou 1984)。

在PCA中,相似性矩阵特征值的获得是按着重要程度降低的顺序,这样对应PCA的轴就依次代表这个矩阵变差由大到小的顺序。因此SUs将要被放置的前几个PCA轴,代表可以被解释的最大变差的百分比(见Gauch 1982的讨论),结果是产生了一个可以提供SUs间生态相似性信息的坐标系统。

在前一章里,我们提出了在群落生态学中与PCA使用有关的问题,是由这种事实产生的,即PCA是线性模型,也就是,一个SU在PCA轴系统空间的坐标,由加权的种多度线性组合确定。如果非线性关系存在,生态数据经常是这样一种情况,一个线性排序模型如PCA将不能很好地提供SUs间真正的关系。Orloci定义这种情况为A型畸变,在第21章里,我们讨论一些发展了的非线性排序模型,以减少这种畸变。Gauch和Whittaker(1972)和Noy-Meir和Whittaker(1977)建议PCA最好是应用在假定线性模型可以使用,即环境和组成变化相对狭窄范围内获得的数据上。

19.2 步骤

我们下面概括的PCA方法采用的是Austin和Orloci(1966)和Orloci(1966, 1967b, 1973,

1978)所描述的步骤,PCA 可用众多的方法来计算,我们认为 Orloci 的方案是最适合教学法的,他的方案如图 19.1 所示,详细的计算步骤如下,我们鼓励学生通过这节的步骤和 19.3 节的内容(例子计算),同时逐步地研究。

步骤 1. 数据矩阵标准化。 作为个提醒,回忆一下群落数据矩阵 X (图 17.1) 有 S 行(种)和 N 列(SUs),这个用矩阵的记法可以写成 $X_{S \times N}$, x_{ij} 是第 i 行(种),第 j 列(SU)观测值(如密度,生物量),Orloci 方案中的第一步是对这个 X 矩阵的标准化,即每个 x_{ij} 元素用下式代替

$$a_{ij} = \frac{x_{ij} - \bar{x}_i}{F_i} \tag{19.1}$$

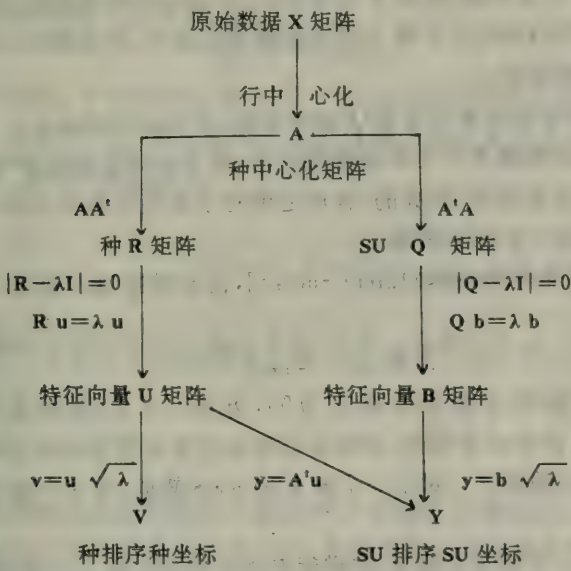


图 19.1 PCA 排序中 R/Q 两用(种和 SU)方案,以 Orloci(1966,1967b,1973,1978)文章为基础

\bar{x}_i 是第 i 行(种)平均值, F_i 是一个标准化函数。在最简单的情形中,如果我们选择 $F_i = 1$, 在公式(19.1)中这就产生一个所谓的行中心化(Row centering),也就是原始坐标系向种空间的形心(重心)的一种移动(见下面的和 Orloci 1974a)。我们这里选择的标准化函数是:

$$F_i = \sqrt{\sum_{j=1}^N (x_{ij} - \bar{x}_i)^2} \tag{19.2}$$

也就是关于行平均数的平方标准差总和的平方根。利用公式(19.1)带有元素 a_{ij} 的一个新矩阵 $A_{S \times N}$ 就形成了。

步骤 2. 计算种和 SU 相似性 R—方式(种)相似性由后乘 A 的转置矩阵获得:

$$R_{S \times S} = A_{S \times N} A'_{N \times S} \tag{19.3}$$

R 是所有种对间皮尔逊积矩相关 (Pearson product-moment correlations) r_{ij} 的矩阵, A_i 是 A 的转置矩阵(行和列转化)。来自公式 19.3 的相关系数是由于我们在公式(19.2)中 F_i 的特定选择。

Q-方式相似由下式给出

$$Q_{N \times N} = A'_{N \times S} A_{S \times N} \quad (19.4)$$

Q 是所有第 i 和第 j 个 SU_s 间的 q_{ij} 数(量)积相似性(Scalar product similarities)矩阵(对于不熟悉矩阵相乘的同学我们建议重新看一下 Searle(1966)和 Pielou(1984)给出的例子,或通过本书 19.3 例子的仔细研究)。

正如上面导出的那样,证明 R 和 Q 矩阵包含基本相同的信息(Noy-Meir 和 Whittaker 1977),因此 SU 排序下面剩余的步骤是使用以上两种任一矩阵去获得相同的结果,这已在图 19.1 中阐述了。在一个样本中 SU_s 的数目(N)如果远远大于种的数目(生态研究中通常的情况),那么 R 方案的计算更为有效(即关于 R 矩阵的 PCA),如果 S 远远大于 N, Q 方案将被选来计算(即关于 Q 矩阵的 PCA),一个 R 方案的步骤描述如下,但从图 19.1,同学们应能看出 Q 方案的数学计算步骤是相同的。

步骤 3. 计算 R 的特征值和特征向量(Eigenvalues 和 eigenvectors)。关于矩阵的特征根和特征向量计算方法的详细描述不在本书范围之内,有许多计算机程序可以进行这方面的分析,另一方面,学生们在依赖这些“现成的”程序之前对于这些计算至少应有个鉴别,这也是十分重要的,因此,下面给出一个总的轮廓。

$R_{S \times S}$ 矩阵的特征值[也称特征根(latent roots)],由希腊字母 λ 表示,可解下面的方程获得:

$$|R_{S \times S} - \lambda I_{S \times S}| = 0 \quad (19.5)$$

I 是个单位矩阵(主对角线元素为 1,其它为 0),方程(19.5)是特征方程(Characteristic equation),是一个关于 λ 的 S 次多项式,两条竖杠表示计算结果矩阵的行列式(不熟悉这些术语的同学可参见 19.3 节中的例子)。对于一个象 R 的 $S \times S$ 矩阵,有 S 个根(即 $\lambda_1, \lambda_2, \dots, \lambda_S$)需要被解出。

与每个 λ_i 有关的是一个特征向量 u_i ,这样就有 S 个特征向量。 u_i 元素的数目(一个列向量)等于 R 的阶数(行和列的数目)。因此, u_i 的维数是 S 行一列,特征向量由下面的恒等式获得。

$$Ru_i = \lambda_i u_i \quad (19.6)$$

方程的解法在 19.5 节讲述,尽管在 R 中有 S 个特征值和特征向量,在实际中生态学家只对前两三个特征值和特征向量感兴趣,因为它们代表了最大数量可以解释的方差,而且希望产生一个有信息的排序。每个特征值占总方差的百分比由与一个特征值除以 R 矩阵的迹(主对角线元素之和)计算。

步骤 4. 调节每个特征向量。特征向量经常被调节(或标准化),这样它们的内积或平方和等于 1,也就是:

$$u_i' u_i = 1 \quad (19.7a)$$

这通过对每个特征向量计算调节因子 k_i 来完成。 k_i 是第 i 个特征向量元素平方和的平方根倒数,也就是:

$$k_i = \frac{1}{\sqrt{\sum_{q=1}^S u_{iq}^2}} \quad (19.7b)$$

u_{iq} 是 i 个向量第 q 行的元素,那么对于每个向量有:

$$u_i = k_i \begin{bmatrix} u_{1,i} \\ u_{2,i} \\ \vdots \\ u_{s,i} \end{bmatrix} \quad (19.7c)$$

R 的标准化向量 **S**, 每一个含有 **S** 个元素, 在矩阵 **U_{s×s}** 中很方便地总结为一列向量。

步骤5 计算每个种排序坐标。每个种与第 **i** 个主分量的关系由下式给出

$$v_i = u_i \sqrt{\lambda_i} \quad (19.8)$$

这些调节了的关系可以作为 **PCA** 排序的坐标来绘图。

用矩阵记法, 上面的关系可以写成

$$V_{s \times s} = U_{s \times s} \Lambda_{s \times s} \quad (19.9)$$

V 的元素是第 **i** 个种与第 **j** 个分量的关系, **Λ** 是一个 λ_i 为主对角线元素, 其它元素为0的矩阵

步骤6. 计算 **SU** 排序坐标。 **SUs** 在前三个主分量坐标位置是由 **A** 的转置矩阵后乘以它的对应特征向量获得, 用矩阵的记法为

$$Y_{N \times 3} = A'_{N \times s} U_{s \times 3} \quad (19.10)$$

Y 的行是 **N** 个 **SUs** 在前三个主分量的坐标(元素 y_{ij} 是第 **i** 个 **SU** 在第 **j** 个主分量的值)。

19.3 举例: 计算

在5个 **SUs** 中两个种的多度将被用来阐述 **PCA** 的实际计算过程, 这些数据在表19.1中给出, 组成群落数据矩阵 **X**, 这种简单的数据使用, 使得我们能够很容易展示出 **PCA** 包含的众多计算, 当然我们提醒学生, 在 **PCA** 中应用大的多维的数据, 尽管我们下面阐述的计算包括矩阵的代数学, 同时经常是很冗长乏味的, 但我们还是鼓励学生经常系统地参照图19.1和19.2节进行研究和学习, 将会得到恰当的 **PCA** 结果。

步骤1 数据矩阵标准化。用公式(19.1)和(19.2)对群落数据矩阵 **X** 进行标准化。对于种1 $\bar{x}_1 = 3$

$$F_1 = \sqrt{(2-3)^2 + (5-3)^2 + \dots + (0-3)^2} = 4.24$$

表19.1 2个种在5个 **SUs** 中的多度数据矩阵

种	SUs					种	
	(1)	(2)	(3)	(4)	(5)	平均数	离差 ^a
(1)	2	5	5	3	0	3.0	18
(2)	0	3	4	2	1	2.0	12

^a: 行平均数平方离差总和。

表19.2 行(种)标准化矩阵 A: (a)行中心化和(b)按公式(19.1)和(19.2)被调节因子除得到的结果

(a)	SU _s					种平均数	调节因子
	种	(1)	(2)	(3)	(4)	(5)	
A=	(1)	-1	+2	+2	0	-3	0.0
	(2)	-2	+1	+2	0	-1	0.0
(b)	$A = \begin{bmatrix} -0.24 & +0.47 & +0.47 & 0.0 & -0.71 \\ -0.63 & +0.32 & +0.63 & 0.0 & -0.32 \end{bmatrix}$						

类似地对于种2

$$F_2 = 3.16$$

A 的新的标准化元素由公式(19.1)计算,例如在 SU5中的种2是

$$a_{2,5} = (1 - 2)/3.16 = -0.32$$

每一行元素通过除以对应于它的离差被标准化,这个离差是对应的关于行平均数的平方离差总和的平方根。完成标准化的 A 矩阵如表19.2所示,注意行平均数是0。

五个 SU_s 位置分别表示在原来的两个种坐标空间和种中心化后的空间,如图19.2a和19.2b所示,注意在种中心化的情况下,SU1落在形心点,因为它的种多度等于行平均数。

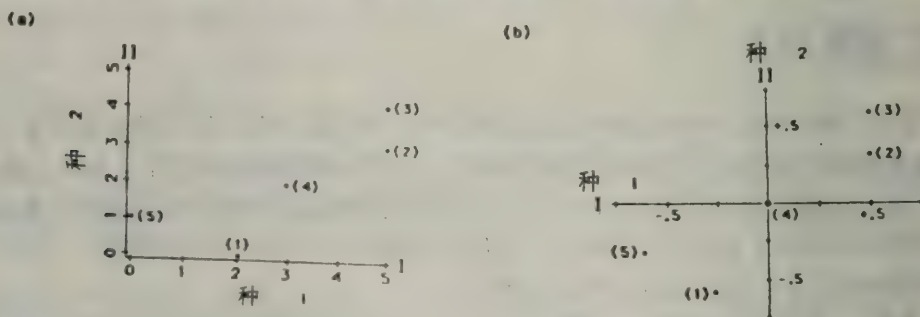


图19.2 五个 SU_s 在(a)两个种空间(b)种一中心化空间的位置

步骤2. 计算种相似性 利用公式(19.3)相关如下。

$$R = \begin{bmatrix} -0.24 & +0.47 & +0.47 & 0.0 & -0.71 \\ -0.63 & +0.32 & +0.63 & +0.0 & -0.32 \end{bmatrix} \begin{bmatrix} -0.24 & -0.63 \\ +0.47 & +0.32 \\ +0.47 & +0.63 \\ 0.0 & 0.0 \\ -0.71 & -0.32 \end{bmatrix}$$

对于第1行,第1列,也就是元素 $r_{1,1}$

$$r_{1,1} = (-0.24)(-0.24) + \dots + (-0.71)(-0.71) = 1.0$$

类似地,

$$r_{1,2} = +0.82; r_{2,1} = +0.82; r_{2,2} = +1.0$$

(注意 $r_{1,2} = r_{2,1}$, 因为 R 是对称的) 以矩阵形式表示如下

$$R = AA' = \begin{bmatrix} 1.0 & +0.82 \\ +0.82 & 1.0 \end{bmatrix}$$

步骤3. 计算 R 的特征值和特征向量。由公式(19.5), 特征方程由下式给出:

$$\begin{aligned} |R - \lambda I| &= \left| \begin{bmatrix} 1.0 & +0.82 \\ +0.82 & 1.0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| \\ &= \left| \begin{bmatrix} 1.0 & +0.82 \\ +0.82 & 1.00 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| \\ &= \left| \begin{bmatrix} 1 - \lambda & +0.82 \\ +0.82 & 1 - \lambda \end{bmatrix} \right| = 0 \end{aligned}$$

一个含有 a, b, c 和 d 四个元素矩阵的行列式计算如下:

$$\left| \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right| = (ad) - (bc)$$

因此, 特征方程可以写成

$$(1 - \lambda)^2 - (0.82)^2 = 0$$

通过重新整理就获得了以 λ 表示的2次多项式。

$$\lambda^2 - 2\lambda + 0.33 = 0$$

应用二次方程, 可以解出这个多项式的两个特征值:

$$\lambda = \frac{-2 \pm \sqrt{(-2)^2 - 4(0.33)}}{2} = \frac{-2 \pm 1.64}{2}$$

所以

$$\lambda_1 = \frac{-2 - 1.64}{2} = \frac{-3.64}{2} = 1.82 (\text{忽略符号})$$

和

$$\lambda_1 = \frac{-2 + 1.64}{2} = \frac{-0.36}{2} = 0.18 (\text{忽略符号})$$

注意这两个特征值的总和与 R 的迹相等, 迹即主对角线元素之和。

每个特征值所占的观测值总方差的百分比是由它作为迹的一个分数来计算:

$$\% \text{ 总方差}(\lambda_1) = \lambda_1 / \sum_{i=1}^S r_{i,i} = 1.82/2.0 = 91\%$$

$$\% \text{ 总方差}(\lambda_2) = \lambda_2 / \sum_{i=1}^S r_{i,i} = 0.18/2.0 = 9\%$$

由于本例计算了所有 S 特征值, 即 100% 方差各有归属, 在实际中, 情况并非如此。

与特征值有关的特征向量由公式 (19.6) 计算, 对于 $\lambda_1 = 1.82$ 的特征向量计算是

$$\begin{bmatrix} 1.0 & +1.82 \\ +1.82 & 1.0 \end{bmatrix} \begin{bmatrix} u_{1,1} \\ u_{2,1} \end{bmatrix} = 1.82 \begin{bmatrix} u_{1,1} \\ u_{2,1} \end{bmatrix}$$

由此获得两个未知的方程

$$1.0u_{1,1} + 0.82u_{2,1} = 1.82u_{1,1}$$

$$0.82u_{1,1} + 1.0u_{2,1} = 1.82u_{2,1}$$

通过重新排列

$$1.0u_{1,1} - 1.82u_{1,1} + 0.82u_{2,1} = 0$$

$$0.82u_{1,1} + 1.0u_{2,1} - 1.82u_{2,1} = 0$$

和

$$-0.82u_{1,1} + 0.82u_{2,1} = 0$$

$$+0.82u_{1,1} - 0.82u_{2,1} = 0$$

解这组方程, 我们人为地使 $u_{1,1} = 1$ 那么可以解 $u_{2,1}$

$$+0.82u_{2,1} = +0.82$$

$$-0.82u_{2,1} = -0.82$$

因此 $u_{2,1}$ 等于 1, 这样特征值 λ_1 的特征向量是

$$u_1 = \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix}$$

类似地与特征值 λ_2 有关的特征向量是

$$u_2 = \begin{bmatrix} -1.0 \\ +1.0 \end{bmatrix}$$

步骤4. 调节每个特征向量。由公式(19.76)对于每个 u_i 的调节因子 k 是

$$k_1 = \frac{1}{\sqrt{1^2 + 1^2}} = 0.707$$

和

$$k_1 = \frac{1}{\sqrt{(-1)^2 + (1)^2}} = 0.707$$

对于每个 u_i , 由公式(19.7c),

$$u_1 = 0.707 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix} ; \quad u_2 = 0.707 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix}$$

注意由公式(19.7)

$$u_1' u_1 = 0.707 \times 0.707 + 0.707 \times 0.707 = 1$$

和

$$u_2' u_2 = (-0.707) \times (-0.707) + 0.707 \times 0.707 = 1$$

在 U 矩阵中这两个特征向量很方便地写成列向量形式:

$$U = \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix}$$

步骤5. 计算种排序坐标。由公式(19.8), 关于第 I 向第 II 主分量种的关系是

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix} \sqrt{1.82} = \begin{bmatrix} 0.95 \\ 0.95 \end{bmatrix} ; \quad v_2 = \begin{bmatrix} -0.707 \\ 0.707 \end{bmatrix} \sqrt{1.82} = \begin{bmatrix} -0.30 \\ 0.30 \end{bmatrix}$$

由公式(19.9), 两者择一

$$V = \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix} \begin{bmatrix} \sqrt{1.82} & 0 \\ 0 & \sqrt{1.82} \end{bmatrix} = \begin{bmatrix} 0.95 & -0.30 \\ 0.95 & 0.30 \end{bmatrix}$$

这些相关总结在表19.3中, 两个种与第一主分量轴都具有较高的正相关, 与 PCA 第 II 轴在轻微(负)相关(图19.3)。

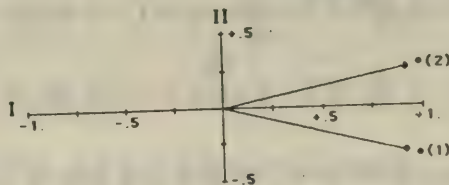


图19.3 种排序显示的种1和种2关于主分量 I 和主分量 II 的相关

表19.3 标准化 A 矩阵的特征分析给出的(a)种坐标(相关)和(b)在两个主分量上 SU 坐标

(a)种坐标		
	主分量	
	I	II
(1)	0.95	-0.30
(2)	0.95	0.30
(b)SU 坐标		
	I	II
(1)	-0.62	-0.28
(2)	0.55	-0.11
(3)	0.78	0.12
(4)	0.00	0.00
(5)	-0.72	0.28

步骤6. 计算 SU 排序坐标。 利用公式(19.10)

$$Y = A'A = \begin{bmatrix} -0.24 & -0.63 \\ +0.47 & +0.32 \\ +0.47 & +0.63 \\ 0.0 & 0.0 \\ -0.71 & -0.32 \end{bmatrix} \begin{bmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{bmatrix}$$

这里对于主分量 I (第1列)SU 坐标是

$$y_1 = (-0.24) \times 0.707 + (-0.63) \times 0.707 = -0.62$$

...

$$y_5 = (-0.71) \times 0.707 + (-0.32) \times 0.707 = -0.72$$

对于主分量 II (第2列)SU 坐标.

$$y_1 = (-0.24) \times (-0.707) + (-0.63) \times 0.707 = -0.28$$

...

$$y_5 = (-0.71) \times (-0.707) + (-0.32) \times 0.707 = 0.28$$

这些 SU 坐标总结在表19.3b 中, SU 排序结果如图19.4所示.将该图同原来的坐标系统相比较(图19.2),并注意两种空间(两个轴系统)是怎样用仅有一点畸变(方差占92%)的单一轴代表(主分量 I)来反映 SU 间的关系,可以很明显地看出:PCA 即使对于非常简单(2个种5

个SU)的计算仍然是很冗长乏味的,因此计算机化的算法是必须的。从一个平方的对称的R一方式和Q一方式矩阵中提取特征值和特征向量的有效方法是很多的。统计软件包如SAS(Ray 1982)和BMDP(Dixon和Brown 1979),包括PCA都是很容易使用的。在试图用PCA处理复杂数据之前,我们提醒学生首先使用已知成分的简单数据(如本例中使用的这个)以便对这些软件包输出的结果熟悉,在许多生态应用中同需要的相比这些“现成的”程序经常产生过多的结果。但从正的一面来看,这些程序经常提供图示输出的选择,这些对于确定SU_s间相似性关系是必要的。

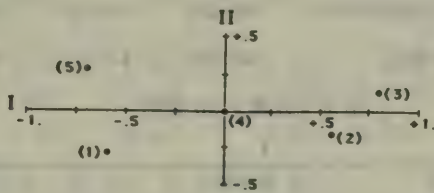


图19.4 在主分量 I 和 II 上5个SU_s的放置

19.4例子：巴拿马蟑螂

利用5种蟑螂在6个SU_s多度数据矩阵,用BASIC程序PCA.BAS(见软盘)进行R方案PCA排序,结果如表19.4所示。

在一个坐标系统内,SU_s的坐标可以用来图示巴拿马6个发生地(SU_s),在此坐标系统内发生地(SU_s)相对位置反映了相似性(图19.5),主分量I占超过60%的方差,发生地3、4、5非常相似,都以5个种的多度较低为特点,发生地6包含单一的种(1),而在发生地1中,种3、4、5的多度较高,为了简单起见,在图19.5中,没有显示种,只画了两个分量,这个排序的进一步解释见24章。

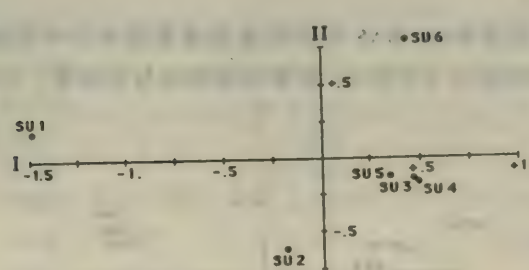


图19.5 巴拿马6个发生地(SU_s)主分量排序

表19.4 利用五种樟脑在6个SU_s中的多度数据进行的主分量分析

R 矩阵(种相关—没有表示)					
R 矩阵的前三个特征值=3.026 1.175 0.790					
特征向量:					
向量1	0.154	-0.255	-0.564	-0.564	-0.543
向量2	0.755	-0.589	0.100	0.192	0.193
向量3	-0.572	-0.708	0.173	0.260	-0.271
每个特征值占追迹百分比			60.5%	23.5%	15.8%
特征值累积百分比			60.5%	84.0%	99.8%

前三个分量上 SU 坐标			
SU	I	II	III
(1)	-1.508	0.206	0.187
(2)	-0.175	-0.660	-0.593
(3)	0.454	-0.112	0.314
(4)	0.477	-0.116	0.307
(5)	0.326	-0.128	0.198
(6)	0.425	0.809	-0.413

在前三个分量上的种相关			
种	I	II	III
(1)	0.268	0.818	-0.509
(2)	-0.443	-0.638	-0.629
(3)	-0.981	0.108	0.154
(4)	-0.950	0.208	0.231
(5)	-0.945	0.209	-0.241

19.5 例子:威斯康星森林

利用 PCA 的 R 方案和 PCA.BAS 程序对威斯康星森林10个立地8个树种的多度数据排序,排序结果在表19.5和图19.6中给出,结果表明栎树与主分量 I 正相关;椴树、铁木、槭

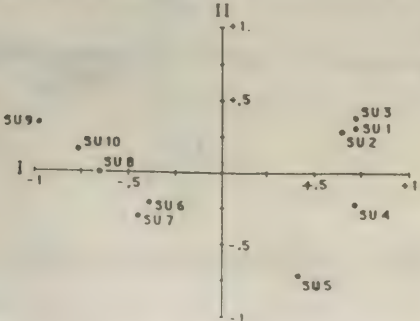


图19.6 南威斯康星10个高地森林立地(SU_s)的主分量排序

树与主分量 I 呈负相关,在主分量 I 这些 SU 以它们种的多度为基础被清楚地分离, SU_s1-4 栎树占优势, SU_s8-10 椴树、铁木、槭树占优势, SU_s5-7 有较多树种的混合。黑栎树和红栎树分别与主分量 I 呈正相关和负相关。因此, SU_s5 由于红栎多度高, 向主分量 I 负的方向移动, 而 SU_s1-3 由于黑栎多度高, 故向主分量 I 正的方向移动。SU_s4 有些黑栎树, 但红栎树更为丰富些, 故向主分量 I 负的方向移动, 这个排序的进一步解释见 12 章。

表 19.5 以 8 个树种多度为基础 10 个高地森林立地 (SU_s) 的主分量分析

R 矩阵(种相关—见表 12.2)									
R 矩阵的前三个特征值=4.472 1.373 1.166									
特征向量:									
向量1	0.378	0.415	0.301	-0.116	0.119	-0.461	-0.387	-0.453	
向量2	-0.021	0.364	-0.409	-0.775	-0.171	0.110	0.186	0.151	
向量3	-0.308	0.101	0.313	0.021	-0.794	0.080	-0.389	0.091	
每个特征值占追迹的百分比	55.9%	17.2%	14.6%						
特征值累积百分比	55.9%	73.1%	87.6%						
前三个分量上 SU 坐标									
SU	I	II	III						
(1)	0.743	0.310	0.051						
(2)	0.691	0.293	0.050						
(3)	0.744	0.369	0.136						
(4)	0.715	-0.263	-0.068						
(5)	0.330	-0.710	-0.338						
(6)	-0.413	-0.438	0.648						
(7)	-0.452	-0.196	-0.535						
(8)	-0.665	-0.065	0.456						
(9)	-0.942	0.386	-0.077						
(10)	-0.751	0.313	-0.322						
前三个分量上的种相关									
种	I	II	III						
(1)	0.798	-0.025	-0.333						
(2)	0.877	-0.427	0.109						
(3)	0.636	-0.479	0.338						
(4)	-0.246	-0.908	-0.023						
(5)	0.253	-0.200	-0.857						
(6)	-0.975	-0.129	0.086						
(7)	-0.818	0.218	-0.420						
(8)	-0.958	0.178	0.098						

19.6 主分量排序的附加讨论

我们介绍的是依照 R 方案进行的 PCA 排序, SU 排序是由种的相关矩阵导出的,而不是由

SU 间的相似性(Q 方案)导出的,在许多生态应用中这将是最有效的方案,由于 R 的秩(种的数目)将典型地小于 Q 的秩(SUs 数目),但如果种的数目超过了 SUs 的数目,应遵循 Q 方案(图 19.1),只要 Q 矩阵是由如 19.2 节描述的那样由种—中心化(行中心化)A 矩阵导出的,排序的结果将是一样的。

Q 矩阵由 SU—中心化(列中心化)A 矩阵,通过转秩这个原始数据矩阵并将它代入“现成”程序的统计软件包中,导出的结果是错误的。用这种作法,PCA 特征值和向量的计算是关于 SU(而不是种)相关矩阵,SU 排序的结果将与我们在这一章里推荐的方案的结果不一样。事实上第一轴将不与最大的变差相一致,这主要是由于排序效能的损失造成的(Orlaci 1966)。

与行中心化相对应的进行 PCA 的运算方法是用非中心化数据矩阵(Carleton 1980, Noy—Meir 1973, Noy—Meir 等 1975)。通过利用非中心化的排序也许能得到一些有用的生态信息。第一主分量在某组数据中获得的变差依赖于整个种—多度格局。这种格局在种中心化过程中被隐含地移走。这些多度格局在生态学研究中也也许是非常有意义的。

另一种方法是在 PCA 计算之前进行双转化(Double transform),双转化包括将数据矩阵的每个元素除以它的行元素总数的平方根和它的列元素总数的平方根(详见 20 章)。通过特征分析计算,导致一个对应分析(Correspondence analysis)形成。

在公式 19.2 中给出了一个计算 A 矩阵元素的调节因子。Goff (1975)进行了一项比较各种 F_i 值对 R 方案排序效果影响的研究。当 $F_i=1$ 时, R 矩阵元素是分布系数。 $F_i=\sqrt{N-1}$, R 矩阵的元素是方差和协方差。一个关于分布和方差协方差矩阵的 PCA 将会给出与相关矩阵类似的结果。不同的仅在度量上,因为只有除数 F_i 的值不同。尽管发现相关系数对于稀有种给出相对大的显著性,但整体结果与分布和协方差的类似,至少对于 Goff 的数据是这样。这再一次指出转化对于分析和进一步的生态解释的重要性。

本章提供的 R 和 Q 方案,种间和 SUs 间的相似性测定是作为数(量)积导出的公式[(19.1)–(19.4)]、Gower (1966)表示将此扩大到 SUs 间欧氏距离是可能的。Orlaci (1973)将此称为 D—方案 PCA,距离矩阵 D 具有的元素由下式给出。

$$D_{N,N} = -0.5[SED_{j,k}] \quad (19.11)$$

$SED_{j,k}$ 是 SUs_j 和 k 欧氏距离的平方(见 14 章 14.2 节),计算 D 的特征值和特征向量,这样就获得了 SUs 的排序,因此进行关于 R、Q、D 矩阵的 PCA 是可能的。

D 方案的扩展称为主坐标分析(Principal coordinates analysis, PCO),这种方法允许利用特征分析减少 D 矩阵的维数以给出 SUs 在欧氏空间的关系。PCO 的详细过程不在本书范围之内,感兴趣的可以参阅 Gower (1966)原始的文章和 Legendre 和 Legendre (1983) Orlaci (1978) 和 Pielou (1984) 的综合处理。

19.7 总结和推荐

1. 主分量作为一种排序方法之所以有吸引力是因为它是一个多元特征分析方法,它通过从 R—方式(种)相似性矩阵或 Q—方式(SU)相似性矩阵获得最大的变差轴产生 SU 排序。同

时,由于 **PCA** 被包括在许多计算机统计软件包之中,是现成可利用的。

2. 对于 **PCA** 产生 **SU** 排序的可能方案中,我们推荐 **Orloci** 的方案,在这种方案里,群落数据是种中心化、标准化的,并且在特征分析之前变换为种间相关矩阵(19.2节),这个方案概念上直截了当,计算上效率高。

3. 尽管对于即使相当简单的群落数据矩阵特征分析的数学计算仍是很冗长乏味的,并且需要一些对矩阵代数的理解,我们还是强烈地希望同学们仔细地过一下这个例子(19.5节),以便在使用现成的程序计算之前对 **PCA** 有个基本的理解。

4. 这个计算例子(19.3节)和蟑螂及森林的例子(分别在19.4,19.5节)阐述了 **PCA** 是怎样“运转”原始数据“种空间”为一个减少维数的新空间(即2个或3个轴的排序),进而显示潜在的群落结构的(见24章)。

5. 如果数据非线性很强,**PCA**(由于是线性模型)将很难代表 **SUs** 间的真正关系,非线性排序模型在21章描述。

6. 当计划用 **PCA** 排序时,我们建议本章所提供的 **PCA** 基本步骤应当同其它的 **PCA** 方案相结合,例如关于非中心化的数据矩阵或 **SUs** 距离矩阵(19.6节),不同方法的结果可用来比较所揭示格局的一致性,如有大的出入,那么你可以调查这种出入是由方法还是数据本身造成的。

第二十章 对应分析

对应分析(Correspondence analysis, COA)和它的变体无偏对应分析(Detrended correspondence analysis DCA)被广泛地应用在排序的技术上,COA 最吸引人的地方在于同时获得对应的 SUs 和种的排序。在一个单一的分析中允许生态学家检验 SUs 和种之间的生态相互关系,COA 可以通过使用特征分析方法获得(同 PCA 类似)或通过一系列加权平均运算[当后面的这种过程被使用时,COA 也经常被称为相互平均(Reciprocal averaging)]。在这一章里,我们使用前一章介绍的特征分析方法来描述 COA,下一章我们将简要描述作为一种对非线性问题排序的方法 DCA。

20.1 基本方法

在南威斯康星森林群落的研究中,Curtis 和 McIntosh(1951)进行了我们第 17 章所称的间接排序首批例子之一。他们根据这些种是演替系列早期或晚期的特征,人为地分配“权”(称为顶级群落适应数目)于森林的树种,这些权从 1(早期演替种,如刺果栎)到 10(顶级种,如糖槭)。以特定立地(SUs)种类组成为基础,获得种多度权数加权平均的得分。对于所有立地使用这种得分就获得了反映演替梯度的一维 SU 排序。这种类型的排序即是指加权平均排序(Weighted average ordination)(Whittaker 1967)。

概念上讲,SUs 排序到一个减少的种空间(图 13.2)简单地包括计算每一个种的加权,之后获得以种加权倍数种多度总和为基础的 SU 得分;因此,SU 的得分或坐标是加权的总和,在这个减少的空间感兴趣的每一维数都可这么作。基本来说,不同排序过程的不同仅在于怎样获得种加权和 SU 得分(Pielou 1984),回忆一下第十九章,PCA 中 SUs 在坐标轴上(主分量上)的坐标是从种相关矩阵的特征值和向量提取出来的,每一个种特征向量所包含的加权在计算 SU 得分时使用。

当使用特征分析的方法时,COA 可以被认为 PCA 的一种简单变体(Pielou 1984)。但 COA 在两个重要方面与 PCA 不同:(1)原始数据转化的形式,(2)特征向量被计算为 SU 坐标或得分的方式。除了 SU 坐标外,在 COA 中还获得了对应的种坐标或相关,因此同时提供了 SU 和种的排序。种的坐标是 SU 排序得分的平均,反过来也这样,SU 排序坐标是种排序相关的平均(Gauch 1982)。

20.2 步骤

用 COA 计算 SU 和种得分的方案如图 20.1 所示,为同第 19 章保持一致,使用与 Orloci (1966,1967b)使用的同样符号:生态数据矩阵含有 S 种(行)和 N 个 SUs(列),种与种 R 矩阵秩为 S,SU 与 SU Q 矩阵秩为 N;X 矩阵的第 i 和第 j 个元素记为 X_{ij} ;A 矩阵的元素是 a_{ij} 等等。特征分析步骤的详细描述见第 19 章。

步骤 1. 数据矩阵 X 的转化。 1. 首先求通过 N 个 SU $i=1$ 到 S 种每个种多度的总和 (即行总和)

$$r_i = \sum_{j=1}^N x_{ij} \quad (20.1)$$

2. 求 $j=1$ 到 N SUs 每个 SU 落在 S 种的多度总和 (即列总和)

$$c_j = \sum_{i=1}^S x_{ij} \quad (20.2)$$

3. 利用行总和或列总和相加计算总的多度。

$$T = \sum_{i=1}^S r_i = \sum_{j=1}^N c_j \quad (20.3)$$

总的多度 T 以后将被用来转化特征向量为 SU 和种关于 COA 的得分。最后, 将第 i 个种在第 j 个 SU 的多度除以第 i 行总和 (r_i) 的平方根和第 j 列总和 (c_j) 的平方根:

$$a_{ij} = \frac{x_{ij}}{\sqrt{r_i} \sqrt{c_j}} \quad (20.4)$$

因此, 利用公式 (20.4), 通过对数据矩阵 $X_{S \times N}$ 双转化形成矩阵 $A_{S \times N}$ (见图 20.1)

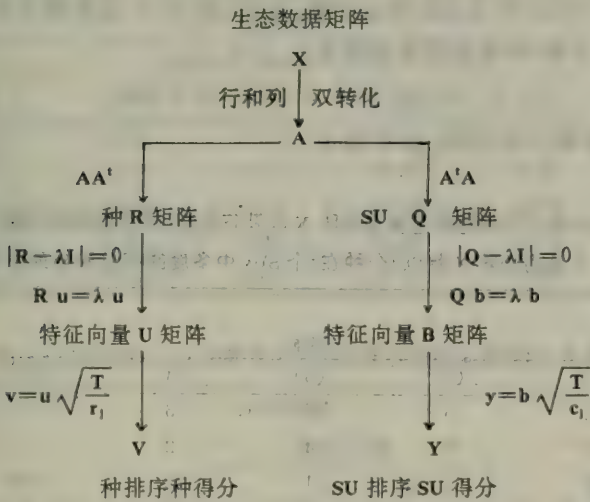


图20.1 同时获得 SU 和种排序的对应分析方案

步骤2. 计算种 (R) 和 SU (Q) 相似性矩阵。 这一步 (图20.1) 同 PCA 的第2步是一致的, 详见 19.2 节, 注意尽管这两步在特征上相似, 但这里由 A 矩阵产生的 R 、 Q 矩阵是不同于 PCA 的 R 、 Q 矩阵的, 因为产生 A 矩阵的转化形式是不同的。

步骤3. 计算 R 和 Q 的特征值和特征向量。 计算平方矩阵的特征值和向量的步骤与 19.2 节 PCA 的步骤也一致, 由双转化而来的 Q 和 R 矩阵的一个独特的特点是特征值一致。

步骤4. 计算种和 SU 排序。 在 COA 成分上的种相关或坐标 v 通过调整特征向量 u 计算, 即:

$$v_i = u_i \sqrt{\frac{T}{r_i}} \quad (20.5)$$

在 COA 分量上 SUs 得分或坐标 y 通过调节特征向量 b 计算即:

$$y_i = b_j \sqrt{\frac{T}{c_j}} \quad (20.6)$$

调节后的种向量 v 以列向量的形式总结在 V 矩阵中, 调节后的 SU 向量 y 以列向量的形式总结在 Y 矩阵中(见图20.1)。为易于解释起见, 这些结果可以绘制成图。

20.3 举例: 计算

我们将用3个种在5个 SUs(表20.1a)中的多度数据来阐述一个完整的 COA 计算过程。之所以说是“完整的 COA”, 是因为按着20.1的方案, 人们可以仅仅选取计算一个 SU 排序的 SU 得分, 而不对种排序作“对应”计算。在此, 我们将对种和 SU 坐标都作一阐述, 而不提供冗长而乏味的特征值和特征向量的计算, 因为这样与19章的内容相重复。

步骤1. 数据矩阵 X 的转化。对于表20.1a 中的数据矩阵, 首先获得所需要的种(行)和 SU(列)总和[公式(20.1)和(20.2)]以及总计数[公式(20.3)], 之后按公式(20.4)转化每个种多度数值。例如在 SU1 中种1转化后的多度值是:

$$a_{1,1} = 2/(\sqrt{15} \times \sqrt{4}) = 0.258$$

行和列(双)转化了的矩阵 A 在表20.1b 中给出

表20.1 (a)在5个 SUs 中三个种的群落多度矩阵 X 以及行总数、列总数和总计数, (b)转化的群落数据矩阵 A 是由矩阵 X 除以3个种在5个 SUs 中多度的行总和与列总和的平方根而来

(a)						
种	SUs					行总和
	(1)	(2)	(3)	(4)	(5)	
(1)	2	5	5	3	0	15
X=(2)	0	3	4	2	1	10
(3)	2	0	1	2	0	5
列总和	4	8	10	5	3	T=30

(b)						
种	SUs					
	(1)	(2)	(3)	(4)	(5)	
(1)	0.258	0.456	0.408	0.346	0.000	
A=(2)	0.000	0.335	0.400	0.283	0.183	
(3)	0.447	0.000	0.141	0.000	0.536	

步骤2. 计算种(R)和 SU(Q)相似性矩阵。种数积 R 矩阵通过 A 后乘它的转置矩阵而得, SU 内积 Q 矩阵通过 A 前乘它的转置矩阵而得(见图20.1), R 和 Q 矩阵在表20.2中给出。

步骤3. 计算 R 和 Q 特征值和特征向量。 R 的三个特征值是:

$\lambda_1=1.00, \lambda_1=0.384, \lambda_1=0.050$

Q 的三个特征值是:

$\lambda_1=1.00, \lambda_1=0.384, \lambda_1=0.050$

正如预料的那样, 它们与 R 的特征值相同。与 R 和 Q 每一个特征值相关的特征向量分别在矩阵 V 和 B 作为列向量给出(表20.3)。

步骤4. 计算种和 SU 排序。在 COA 分量上的种坐标(相关)的计算是通过用 U 的每一个列向量乘以 $\sqrt{\frac{T}{r_i}}$ [公式(20.5)] 所得, 例如种3在分量 III 上的相关是

$v_{3,3} = 0.022 \sqrt{30/5} = 0.054$

表20.2 以 A 矩阵为基础种(R-方式)和 SU(Q-方式)数积相似性矩阵(见表20.1b)

(a)对于种 (R-方式)						
	种	(1)	(2)	(3)		
R=	(1)	0.56	0.41	0.17		
	(2)	0.41	0.39	0.15		
	(3)	0.17	0.15	0.49		
(b)对于 SU _s (Q-方式)						
	SU _s	(1)	(2)	(3)	(4)	(5)
Q=	(1)	0.27	0.12	0.17	0.09	0.23
	(2)	0.12	0.32	0.32	0.25	0.06
	(3)	0.17	0.32	0.35	0.25	0.15
	(4)	0.09	0.25	0.25	0.20	0.05
	(5)	0.23	0.06	0.15	0.05	0.30

表20.3 与(a)R 矩阵和(b)Q 矩阵有关的特征值的特征向量矩阵(见表20.2)

(a)R 矩阵				
特征向量				
种	I	II	III	
U=	(1)	0.707	-0.331	0.625
	(2)	0.577	-0.240	-0.781
	(3)	0.408	-0.913	0.022
(b) Q 矩阵				
特征向量				
SUs	I	II	III	
B=	(1)	0.365	-0.520	-0.765
	(2)	0.516	0.374	-0.104
	(3)	0.577	0.165	0.242
	(4)	0.408	0.294	0.020
	(5)	0.316	-0.690	0.587

表20.4 分别由 U 和 B 矩阵导出的在 COA 三个分量上(a)种相关, V,和(b)SU_s 坐标, Y 的矩阵

(a)种相关, V			
种	COA 分量		
	I	II	III
V=	(1) 1.00	-0.469	0.883
	(2) 1.00	-0.415	-1.35
	(3) 1.00	2.24	0.054

(b)SU 坐标, Y			
SU _s	COA 分量		
	I	II	III
Y=	(1) 1.00	-1.42	-2.10
	(2) 1.00	0.723	-0.202
	(3) 1.00	0.285	0.419
	(4) 1.00	0.721	0.048
	(5) 1.00	-2.18	1.86

在 COA 分量上 SU_s 坐标是通过 B 的每一列向量乘以 $\sqrt{T/c_i}$ 而得, 例如 SU5 在分量 I 的坐标.

$$y_{5,2} = -0.690 \sqrt{30/3} = -2.18$$

种和 SU 在 COA 前三个轴上的坐标在表20.4中给出, 注意在 COA 第一分量上的种相关总是 1, 这是由于双转化和后来的再调节的结果, 当利用另一种方法计算 COA 时, 如人们所知道的相互平均(Reciprocal averaging)(legendre 和 legendre 1983, Pielou 1984)将会产生同样的成分, 这样在构造种的排序时就可以忽视第一个 COA 分量了, 与种坐标情况一样, 在 COA 的第一个分量上 SU_s 的坐标也为 1, 所以也被忽略了。SU_s 和 COA 排序(图20.2)与 PCA 排序的结果是非常相似的(图19.4), 所不同是在 PCA 中仅用了两个种。

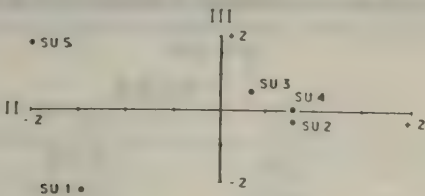


图20.2 . 5个SU 在分量 II 和分量 III 的对应分析排序

20.4 例子:巴拿马蟑螂

应用巴拿马蟑螂的数据矩阵(表11.4a)和 BASIC 程序 COA.BAS(软盘内), 进行种和 SU_s 的 COA 排序, 结果在表20.5中给出, SU 排序如图20.3所示。在 COA 中, SU 的格局仅仅在某些地方可以与 PCA 的结果(图19.5)相比较, 我们把这个留给学生们, 让他们试着解释不同的地方。

表20.5 (a)关于巴拿马6个发生地(SUs)和(b)5种蟑螂的对应分析结果

(a)		SUs 间数积(Q-方式)相似性					
	SUs	(1)	(2)	(3)	(4)	(5)	(6)
Q =	(1)	0.58	0.30	0.07	0.02	0.17	0.23
	(2)	0.30	0.45	0.06	0.08	0.15	0.12
	(3)	0.07	0.06	0.02	0.01	0.02	0.00
	(4)	0.02	0.08	0.01	0.02	0.02	0.00
	(5)	0.17	0.15	0.02	0.02	0.08	0.00
	(6)	0.23	0.12	0.00	0.00	0.00	0.22

Q 的特征值(前三个)

$$\lambda_1 = 1.000 \quad \lambda_2 = 0.239 \quad \lambda_3 = 0.115$$

Q 的特征向量(前三个)

	SUs	(1)	(2)	(3)	(4)	(5)	(6)
向量	1:	0.710	0.553	0.092	0.065	0.263	0.328
	2:	-0.470	0.764	0.093	0.228	0.029	-0.366
	3:	0.492	-0.111	0.260	0.022	-0.191	-0.801

在前三个分量上 SU 坐标

	SUs	(1)	(2)	(3)	(4)	(5)	(6)
分量	I:	1.00	1.00	1.00	1.00	1.00	1.00
	II:	-0.66	1.38	1.00	3.47	0.11	-1.11
	III:	0.69	-0.20	2.80	0.34	-0.73	-2.44

(b) 种间数积(R-方式)相似性

	种	(1)	(2)	(3)	(4)	(5)
R =	(1)	0.04	0.00	0.00	0.00	0.08
	(2)	0.41	0.42	0.14	0.04	0.30
	(3)	0.17	0.15	0.22	0.11	0.28
	(4)	0.00	0.04	0.11	0.06	0.13
	(5)	0.08	0.30	0.28	0.13	0.62

R 特征值(前三个)

$$\lambda_1 = 1.000 \quad \lambda_2 = 0.239 \quad \lambda_3 = 0.115$$

R 的特征向量(前三个)

	种	(1)	(2)	(3)	(4)	(5)
向量	1:	0.065	0.500	0.383	0.174	0.754
	2:	-0.150	0.849	-0.274	-0.235	-0.356
	3:	-0.471	0.057	0.684	0.354	-0.426

在 COA 前三个分量上种坐标

	种	(1)	(2)	(3)	(4)	(5)
分量	I:	1.00	1.00	1.00	1.00	1.00
	II:	-2.29	1.70	-0.72	-1.36	-0.47
	III:	-7.18	0.12	1.79	2.04	-0.56

20.5 例子:威斯康星森林

威斯康星高地森林的数据(表11.6a)COA 结果见表20.6,这些结果是应用 BASIC 程序获得的, SUs 在 COA 中的位置(图20.4)与 PCA 对这些数据处理的结果是非常相似的(图19.6),

在 COA 中主要的不同是 SU8、10与 SU9分离了, 由表11. 6a 可以看出, 在 SU8、9、10中椴树、铁木和糖槭都具有较高的多度, 但 SU8受栎树的强烈影响。

表20.6 (a)10个森林立地(SUs)和(b)8个树种的对应分析结果

(a)										
SUs 间数积(Q-方式)相似性										
SUs	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	0.19	0.18	0.15	0.15	0.11	0.04	0.07	0.03	0.01	0.02
(2)	0.18	0.19	0.15	0.16	0.10	0.04	0.07	0.03	0.02	0.02
(3)	0.15	0.15	0.18	0.16	0.09	0.05	0.06	0.04	0.01	0.04
(4)	0.15	0.16	0.16	0.16	0.13	0.07	0.09	0.06	0.03	0.05
(5)	0.11	0.10	0.09	0.13	0.16	0.10	0.12	0.07	0.05	0.07
(6)	0.04	0.04	0.05	0.07	0.10	0.16	0.11	0.15	0.12	0.12
(7)	0.07	0.07	0.06	0.09	0.12	0.11	0.17	0.14	0.16	0.15
(8)	0.03	0.03	0.04	0.06	0.07	0.15	0.14	0.17	0.16	0.16
(9)	0.01	0.02	0.01	0.03	0.05	0.12	0.16	0.16	0.20	0.18
(10)	0.02	0.02	0.04	0.05	0.07	0.12	0.15	0.16	0.18	0.19
Q 的特征值(前三个)										
$\lambda_1=1.00 \quad \lambda_2=0.54 \quad \lambda_3=0.096$										
SUs	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1:	0.30	0.30	0.29	0.33	0.32	0.30	0.36	0.32	0.30	0.32
2:	0.39	0.39	0.35	0.30	0.11	-0.20	-0.18	-0.31	-0.42	-0.36
3:	0.21	0.24	0.18	-0.14	-0.58	-0.54	0.08	-0.06	0.35	0.29
在前三个分量上 SU 坐标										
SUs	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
I:	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
II:	1.28	1.27	1.22	0.91	0.35	-0.66	-0.51	-0.98	-1.38	-1.11
III:	0.69	0.80	0.64	-0.41	-1.82	-1.78	0.21	-0.20	1.15	0.90
(b)										
种间数积(R-方式)相似性										
种	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
(1)	0.23	0.22	0.17	0.14	0.13	0.03	0.03	0.02		
(2)	0.22	0.30	0.19	0.09	0.12	0.00	0.00	0.00		
(3)	0.17	0.19	0.24	0.18	0.14	0.11	0.05	0.09		
(4)	0.14	0.09	0.18	0.21	0.13	0.16	0.10	0.14		
(5)	0.13	0.12	0.14	0.13	0.15	0.08	0.08	0.08		
(6)	0.03	0.00	0.11	0.16	0.08	0.21	0.16	0.22		
(7)	0.03	0.00	0.05	0.10	0.08	0.16	0.18	0.19		
(8)	0.02	0.00	0.09	0.14	0.08	0.22	0.19	0.25		
R 的特征值(前三个)										
$\lambda_1=1.00 \quad \lambda_2=0.54 \quad \lambda_3=0.096$										
R 的特征向量(前三个)										
种	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
1:	0.35	0.33	0.43	0.41	0.33	0.34	0.27	0.34		
2:	-0.39	0.53	0.20	-0.07	-0.08	-0.40	-0.36	-0.47		
3:	0.08	0.50	-0.39	-0.54	0.02	-0.08	0.48	0.25		
在前三个分量上种坐标										
种	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
I:	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00		
II:	1.11	1.61	0.47	-0.17	0.25	-1.17	-1.32	-1.37		
III:	0.24	1.51	-0.91	-1.30	0.05	-0.24	1.77	0.73		

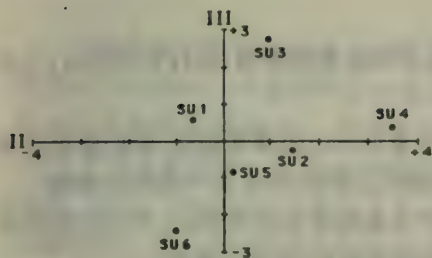


图20.3 巴拿马6个地点对应分析排序

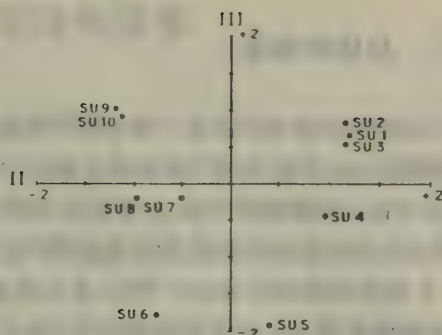


图20.4 南威斯康星10个森林立地(SUs)对应分析排序

COA 中, SU6 与 SU7 也分离开了, 从表 11. 6a 也可以看出, SU6 中种 3、4、5、6 和 8 具有优势, 但 SU7 除了具有以上种外, 种 1、7 也有优势, 因此 SU7 在这个 COA 的排序中被不同地加权。总体上来看, 同 PCA 相比, COA 看上去给出的 SUs 间关系的界线比较明显, 至少对于这些数据是这样。

20.6 对应分析的附加讨论

COA 使用双转化方法代替 PCA 中使用的中心—标准化 (Centred—normalized transformation), 它也可以不同地调节特征向量, COA 的一个主要特点是同时获得 SUs 和种坐标, 这样就构造了“对应的”SU 和种的排序。SUs 和种的坐标是以最大化它们的相互关系的方式而获得。COA 排序应该是最适合排序理论的, 也就是说, 在一个坐标系统内 SUs 和种的位置最好地反映了它们的相似性格局, 作为一种排序方法, COA 的生态效果在 Gauch (1982) 所提供的例子中很明确地得到了表示, 而且 COA 也已成为一个很普遍的排序方法 (见表 17. 1)。

在评述 COA 与极点排序 (PO) 的关系时, Beals (1984) 得出这样的结论: COA “仅有一个轴是基本好的”, 文献来源于“第二轴趋向于弓形, 它的第一轴后面的轴很难解释”, 正如将要在下一章叙述的那样, 弓形的问题是在 PCA 和 COA 排序中 SUs 的结构是明显畸变的, 这是由于种的非线性造成的, 当通过一个较大的环境梯度取样时, 这种现象经常可以观察到, 也就是群落数据具有较高的 β 多样性时 (见第 8 章)。尽管同 PCA 相比, COA 对非线性群落数据不太敏感, 不过 COA 还是有个“弓形”问题 (Gauch 1982), 在处理非线性数据时, COA 的效力可以通无偏处理 (Detrending) 得到增进。

将 COA 同其它的排序方法相比较研究的结论是混杂的。Gauch 等 (1977) 建议同 COA 相比, PO 更优越一些, Whittaker 和 Gauch (1973) 的结论同上面的相反。这种不同与当群落数据有一个高的 β 多样性优势种梯度时用 COA 效果较好有关 (Beals 1984), Beals 还表示称 COA 较好的争论是因为 SUs 被公平地分布于第 I 轴上, 事实上这是个缺点, 因为这个特点也许掩盖了在梯度潜在的第一轴上的真正不连续体。我们支持 Beals 的结论, 没有一个“毫不含糊的”最好的排序方法, 所以同以前几章一样我们推荐对于给定的数据, 同时使用不同的排序方法, 以

检验结果的一致性。

20.7 总结和推荐

1. 对应分析(COA)是一种加权平均技术,它相互地双转化群落数据(通过种和 **SUs**)之后采用特征分析产生“对应”种和 **SUs** 排序(20.1节)

2. COA 的步骤和计算(分别在20.2和20.3节)同主分量(PCA)的非常相似,我们建议学生们对这些进行仔细地研究,以便在使用“现成的”计算机程序之前对 COA 有个基本的理解。

3. 例子表明(20.4和20.5节),在某些情况下,COA 产生的排序格局同 PCA 产生的非常相似(威斯康星森林),但在其它情况下,COA 又给出相当不同的结论(巴拿马蟑螂)。

4. 比较研究表明,当群落有一个优势种和相当广阔的环境梯度时,COA 完成的排序要比 PO 和 PCA 完成的要好。COA 第二轴经常有由非线性种沿某种梯度反映而引起的“弓”形问题(参见下章),而且这种弓形问题经常难以作出生态解释。

5. 由于 COA、PCA 和 PO 方法处理群落数据(并不是它们的目的)的方式不同,我们建议这些方法以及其它的排序方法处理同一组数据时一起使用,任何不同都可以用来洞察存在的群落格局。

第二十一章 非线性排序

群落生态学家经常通过一个较大的环境梯度收集数据,在这种梯度中常见情况是出现在梯度对立两端 SUs 中的种较少或没有。很容易证明出现在这种梯度的个体种—多度格局将是典型的非线性性(如钟形),而且这些非线性多度可以引起 SUs 在种的多维空间分布是弓形(或更严重地螺旋形),这种非线性性/弓形问题对于前面几章所描述的排序方法是一个严重的限制。对这一问题有两个解决方法:(1)利用移走(使无偏)这种 SUs 弓形问题的排序方法,这种排序适合于由中等程度种的非线性性引起的弓形或马蹄形问题;(2)在具有较强的种非线性情况下,使用一种调整 SU 螺旋形的排序方法。在这一章里,我们描述多项式排序即无偏(Detrended),非度量多维调节,它调整排列顺序。

21.1 基本方法

当我们用象主分量分析一类的排序方法分析群落数据时,对于 SUs 观察到的一个典型格局就是“弓形”(如图 19.6)。这种畸变经常给生态解释带来困难。

经常使用的许多排序技术,如极点排序(PO,18 章),主分量分析(PCA,19 章)和对应分析(COA,20 章)是假定 SUs 沿着一个环境梯度取样时种反映出的关系是线性的或接近线性的。这个假定在当潜在的环境梯度狭窄或种的生态幅较宽时是有道理的(Noy-Meir 和 Whittaker 1977)。举个例子,8 个 SUs 沿一假设的梯度取样时,种 A 和 B 是接近于线性关系的,如图 21.1a 所示。在这种情况下,PCA 的第一轴(I)(或一个 COA)对于对应 SU 的格局的配合将是很好的,这样轴 I 将准确地反映环境梯度(图 21.1b)。

如果环境的梯度较宽,但不是绝对的,种 A 和种 B 典型的多度格局将会表现出沿着这个梯度它们个体钟形分布曲线呈部分的重叠(图 21.1c)。在这种情况下,这两个种多度的关系将是明显的非线性性。对应 SU 的格局将是弓形的曲线,如图 21.1d 所示。PCA 的前两个轴将反映这种弓形,因此,在中心化和旋转的坐标系统内就需要用两个轴来代表一个潜在的梯度。如果这种弓形趋势可以通过形成一反映单一环境梯度的单一轴移走(或无偏)。这将使生态解释变得容易。

Phillips(1978)描述了一种使用二次多项式回归展开或无偏化由 PCA 对于中等非线性数据处理得到的弓形曲线,他称这个步骤为多项式排序,我们称之为无偏主分量(Detrended principal component DPC)。对于克服 PCA 的弓形问题,Phillips 的方法是一个相对简便的步骤。在 21.2.1 节我们将描述 DPC。

我们在前一章描述的 COA 也存在这个弓形的问题(如图 20.4),Hill 和 Gauch(1980)描述了一种无偏对应分析(Detrended correspondence analysis,DCA),用于对由 COA 经常可以观测到的 SU 弓形进行无偏化处理。DCA 中使用了一个复杂的算法来去掉弓形(Austin 1985),所以在这章里,我们选择地提供了一个较简单的 DPC 过程,但我们建议使用 COA 的同学如发现有明显的弓形也应考虑使用 DCA。

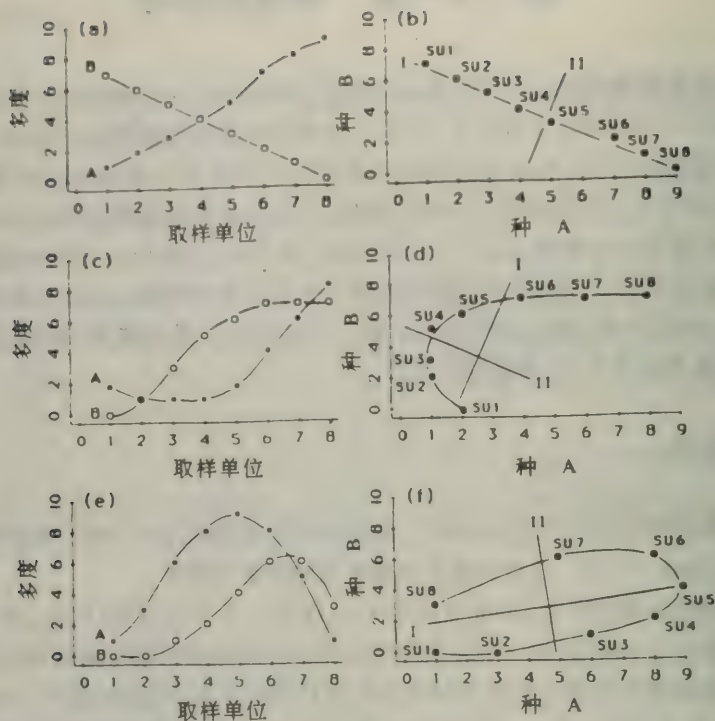


图 21.1 种 A 和 B 在 8 个 SUs 位于单一环境梯度一条样线上以固定间隔放置的多度。(a) 狭窄梯度或宽幅度种典型的多度格局, (b) 在种空间的 SUs 和在这个空间上对 PCA 轴 I 和 II 的配合, (c) 中等梯度或幅度种典型的多度格局, (d) 在种空间的 SUs 和在这个空间上对 PCA 轴 I 和 II 的配合, (e) 宽梯度或窄幅度种的典型多度格局和 (f) 在种空间的 SUs 和在这个空间上对 PCA 轴 I 和 II 的配合

如果潜在的环境梯度非常宽(如从山底到山顶的延伸),通过这么宽梯度种的多度将很可能更加完全地重叠(如图 21.1e)。这种钟形重叠的种曲线导致 SU 格局看上去象个“圈形”(Pielou 1984)或“螺旋形”(Noy-meir 和 Austin 1970)。一个 PCA 方法对于这样强的非线性数据是不能令人满意的(见图 21.1f),因为 PCA 不能打开这样的圈或螺旋以反映单一潜在的梯度。事实上,DPC 也可能不能令人满意。对于这样尖锐的非线性数据结构,各种非线性排序方法,如连续体分析(Continuity analysis)(Noy-Meir 1974),高斯排序法(Gauch 等 1974)或非度量多维调节(NMDS)(Anderson 1971)必须使用。在 21.2.2 节,我们将描述同其它的经常使用的排序方法相比更为有效恢复非线性数据结构的 NMDS 方法。(Kenkel 和 Orloci 1986; Minchin 1984)。

21.2 步骤

21.2.1 无偏主分量(DPC)

下面的 DPC 方法,基本上是 McDonald(1962,1967)描述的非线性因子分析的改写和更名,他把它用到心里测验数据上(Phillips 1978),我们之所以阐述作为简单的非线性排序步骤的 DPC,是因为我们觉得在许多群落生态研究中,中等的非线性数据结构是普遍的;一些众多的例子支持这一点(见 Gauch 1982),同时,DPC 检测非线性性逻辑上应接着 PCA 进行,也就是说,如果有个弓形 SU 格局存在,它作为中等非线性种多度关系的结果,那么 DPC 过程可以检验在 PCA 内产生的弓形 SUs 的存在和显著性。如果显著,就无偏化这个弓形到一个轴上,以反映潜在的环境梯度。

步骤 1. 用 PCA 处理数据(19 章)。这个 PCA 的结果是以从 R 矩阵的特征值和特征向量而来的 SU 坐标为基础的 SU 排序(见 19.2 节)。

步骤 2. 回归 PCA 的成分 I 到成分 I 上。 计算主分量 I 回归到主分量 I 的抛物线(二次)多项式。多项式回归方程是

$$Y = B_0 + B_1X + B_2X^2 \quad (21.1)$$

Y 是在 PC I 上 SU 坐标(因变量),X 是在 PC I 上 SU 坐标(自变量),B 是决定抛物线形状的参数,B₀是截距或位置参数。B₁、B₂ 决定抛物线的开度和方向(如果 B₂ 是正的,抛物线的开口向上,如果 B₂ 是负的,抛物线的开口向下)。在公式(21.1)中的 B 由用多元线性回归估计。这些计算我们建议学生参见 Sokal 和 Rohlf(1981,16.1 节)

步骤 3. 检验回归的显著性。 利用多元测定系数检验抛物线回归的显著性,用 R²(见 Sokal 和 Rohlf 1981,16.2 节)估测 F 比率:

$$F = \frac{R^2/2}{(1-R^2)(N-2-1)} \quad (21.2)$$

N 是 SUs 数目,如果抛物线回归是显著的(P=0.05 df=2,N-2-1),那么原来的 PCA 轴 I 和轴 I 就可以结合为一单一的无偏主分量轴,这个单一的 DPC 轴应该能较好地反映产生 PCA 弓形潜在的环境梯度。

步骤 4. 构造 DPC 轴。 通过对在二维的 PCA 系统的 SUs 位置到抛物线的垂直投影,一个单一的 DPC 轴就构造成了。通过目视结果,这很容易在图上做到。使位于端点的 SUs 作为轴的端点,人为地使 DPC 在-1 到+1 范围内(见 21.3.1 节),这种方法对于简单的数据表现的很好,Phillips(1978)提供了关于复杂数据和更精确的代数方法。这个抛物线就这样通过形成一单一的 DPC 排序而被无偏化。

21.2.2 非度量多维调节(NMDS)

下面的 NMDS 排序步骤是以 Fasham(1977),Pimental(1979)和 Prentice(1977)的描述和在心里测验学上的发展(Shepard 1980)为基础的。在那些种间具有很强的非线性关系和在多维空间 SUs 结果呈圆形或螺旋形的情况下(如图 21.1f),NMDS 是一个有效的排序方法,NMDS 的一些形式看上去要比其它的方法有优势。

步骤 1. SUs 对的顺序排列。首先计算 Q -方式所有 SUs 间的相似性(如 14 章中的任何一种距离系数)。之后这些测量值以上升的顺序也就是从最低到最高的相异性(ds)按下面形式排列。

$$ds_{i,j} < ds_{k,l} \quad (21.3)$$

SUs 对 $i < j$ 和 $k < l$ 。Kenkel 和 Orloci(1986)使用模拟的数据比较了排序方法,并发现最健全的 NMDS 方案是首先由 SU 标准化数据;例如弦距离[公式 14.9]就可以作这个。

因为 NMDS 在以后的步骤中仅仅使用非度量排列顺序(而不是实际的 Q -方式测量),故在 NMDS 中排列 SUs 所选择的相似性测量方法常常被指出是不重要的。NMDS 对于所选择的方法没有其它排序方法那么敏感,关键是使用的方法应允许用一个调节的方法(象 NMDS)沿着一条直线反映潜在的生态梯度去正确地排列 SUs 。这是一个相当复杂的问题,我们将在 21.6 节进一步讨论。

步骤 2. 顺序排列 SUs 的排序距离。使用前面的两维或三维排序 SUs 的坐标,我们计算所有对 SUs 间的距离,之后这些距离以上升的顺序排列,也就是:

$$d_{i,j} < d_{k,l} \quad (21.4)$$

SUs 间 $i < j, k < l$ 。14 章中的任何距离方法都可以利用。事实上,这种方法不需要是欧氏的(Fasham 1977)。

这样作的目的就是提出一个最初的布局,这个布局是以一个有道理的排序结果(如 PO 、 PCA 或 COA)为基础的,这样才能使步骤 3 的优化算法能够发现“总体”最小值,而不是某些“局部的”最小(Fasham 1977),为了解释这个难以理解的意思,我们回到第七章与之相类似的问题上,为了估测对数正态模型(公式 7.3)的参数值,我们描述了几种用来提供有道理的最初估计值的方法;很明显,最初的估计值越好,最小化观测的与期望的频度值的反复迭代步骤就越少。这里的问题与之类似,最初的估计值越好(排序的顺序排列),最小化(下面的步骤 3) $SU \times SU$ 距离(步骤 1)与排序距离(步骤 2)的差异所需要的反复迭代步骤的数目就越少。一个差的最初估计值常常导致“部分”最小和一个无效的排序;一个好的最初估计值将会产生一个“总体”的最小。

如果来自群落数据的 SU 相似性排列顺序,与来自排序的 SU 距离完好地匹配(作为一个单调递增函数,步骤 3),那么最初的布局就可以认为是一个令人满意的 NMDS 排序。基本的, NMDS 认为一个较好的排序在于群落数据的 SU 与 SU 间的距离和排序空间之间应有一个最佳的排列顺序关系。但一个完全的匹配是不可能的(除了非常小的数据,21.3 节),所以我们进行第三步。

步骤 3. 优化排列顺序。通过利用 Kruskal 的非参数回归技术首先计算一条单调递增曲线,这条曲线对原始数据 SU 和 SU 相似性 $ds_{i,j}$ 和排序的 SU 和 SU 距离 $d_{i,j}$ 的关系给出一个最佳的配合,这样就建立了一种反复迭代算法。对于每个 SU 对, i 和 j , 每个 $ds_{i,j}$ 都有一个回归估测值 $\hat{d}_{i,j}$, 这些估测值不是距离,而是一组与 $ds_{i,j}$ 是单调关系的数字,用它们来计算背离单调性的程度。Kruskal(1964)定义这种程度为应力(Stress)。

$$SR = \sqrt{\frac{\sum_{i < j}^M (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j}^M (d_{ij}^2)}} \quad (21.5)$$

求和是对所有SU (i,j)对, $i < j, M = [N(N-1)/2]$. SR 值越小, 接近于单调性的关系越密切, 也就是, 较低的 SR 值表明排序的 SU 和 SU 距离与群落数据 SU 和 SU 相似性的排列顺序具有较高的和协程度。

通过反复迭代, 利用 NMDS 算法在排序空间内移动 SU 以便应力(SR)最小, 这个最小化的过程, 称之为“最速下降(Steepest descent)方法”, Kruskal(1964)对其进行了描述, 数学的详细过程不在本书范围内, 但这种过程已建立在我们的 NMDS. BAS 程序中(见软盘)。

步骤 4. 构造 NMDS 排序。在这种布局下经反复迭代后, SR 已不显著, 通过标准化这种布局(即 0 形心化和单位分布)就形成了最终的排序。如果这些 NMDS 轴(象 SU 坐标定义的那样)强烈地相关, 它们可以被旋转为没有相关的主轴(如在 PCA 中, 19 章)旋转结果的轴坐标用来构造最终 NMDS 排序, 这个旋转的选择在程序中提供。

21.3 举例: 计算

21.3.1 无偏主分量(DPC)

为了阐明关于生态数据的 DPC 排序, 将利用 PCA 关于 2 个种在 5 个 SU 中的数据表 19.1 排序的结果。

步骤 1 用 PCA 处理这组数据。表 21.1 中给出了 5 个 SUs 在主分量(PCs) I 和 II 上的得分。

表 21.1 回归变量作为 5 个 SUs 在两个主分量(PC I 和 PC II)上的坐标以及平均数(Mean)、标准差(SD), 相关(r), 回归系数(B's)和测量系数(R²)的估测值

变量	取样单位					Mean	SD
	(1)	(2)	(3)	(4)	(5)		
Y=PC II	-0.28	-0.11	+0.12	-0.50 ^a	+0.28	-0.10	0.31
X=PC I	-0.62	+0.55	+0.78	0.00	-0.72	0.00	0.67
X ² =(PC I) ²	+0.38	+0.30	+0.61	0.00	+0.52	0.36	0.23
$r(Y, X)=0.00, r(Y, X^2)=0.88, \text{ 和 } r(X, X^2)=0.05$							
$B_0=-0.52, B_1=-0.018, B_2=1.17, \text{ 和 } R^2=0.77$							

a. 为了更有效地阐述抛物线回归对于主分量 I 和 II 的配合情况, 我们人为地把 SU 4 放在 0.0 和 0.5 处, 以代替 19 章所表示的 0.0 和 0.0 处, 图 19.4。

步骤 2. 回归 PCA 成分 II (Y) 到成分 I (X) 上。按 Sokal 和 Rohlf(1981, 16.1 和 16.2 节)概括的那样首先计算平均数、标准差、相关系数、回归系数的估测值。表 21.1 提供了这些估测值。回归结果是[公式(21.1)]:

$$Y = -0.52 + (-0.018x) + 1.17x^2$$

步骤 3. 检验回归的显著性。 利用多元测量系数 $R^2=0.77$, 计算 F 比率[公式(21.2)]:

$$F = (0.77/2)/[(1 - 0.77)/(5 - 2 - 1)] = 3.55$$

这个 F 统计量不显著($P=0.05$, $df=2,2$), 在这个例子中, 统计的不显著性是由于自由度较低; 抛物线对数据配合的较好的缘故(见图 21. 2a)

通常情况下, 如果这一步给出了不显著性, 那么我们就可以认为(1)线性 PCA 适当地归纳了这组数据或(2)存在一个强的非线性数据结构, 圆形和螺旋形不能被展开, 但为了阐明计算过程, 我们假定回归是显著的并进行第四步。

步骤 4. 构造 DPC 轴。 通过作 SUs 到抛物线的垂直投影, 就构造了 DPC 轴(图 21. 2a), 在 DPC 轴上, 两端 SUs 和 SUs 初分别定为 -1 和 $+1$, 其它的 SUs 通过“目视”它们之间的相对位置也被投影到这个轴上, 这样 5 个 SUs 在 $PC\ I$ 和 $PC\ II$ 上弯曲的格局就被单一的展开的 DPC 轴代替。

21. 3. 2 非度量多维调节

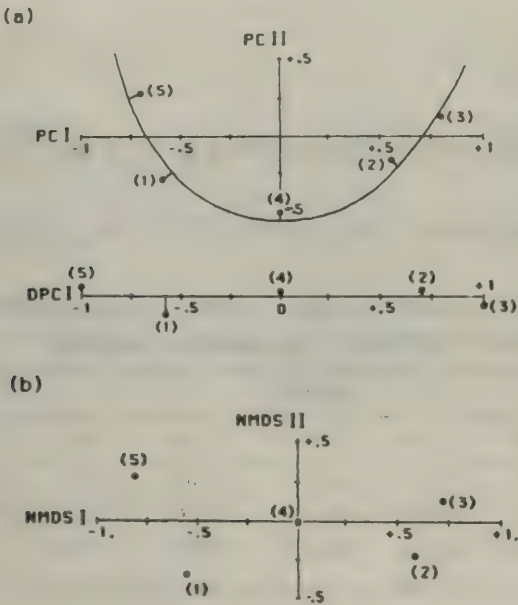


图 21.2 5 个 SUs 的排序 (a)主分量分析表示的 SU 在分量 I ($PC\ I$) 和 II ($PC\ II$) 上的位置, 并把 $SU\ 4$ 人为地放在 $(0.0, -0.5)$ 处, 和 $PC\ II$ 到 $PC\ I$ 上的二次多项式回归的配合以及无偏主分量轴 I ($DPC\ I$); (b)在两个 NMDS 轴上非度量多维调节表示的 SUs 位置。

我们再一次利用这个 2 个种在 5 个 SUs 中的人为数据(表 19. 1)和由 PCA 对这组数据处理的结果作为一例子。

步骤 1. 顺序排列 SU 对. 对所有 (i,j) SU 对进行 Q—方式相似性测定, 在此利用欧氏距

离[ED, 公式(14.1)], 并按着相异性由低向高排列[公式(21.3)]:

顺序	1	2	3	4	5	6	7	8	9	10
对	$ds_{2,3} < ds_{2,4} \leq ds_{1,5} \leq ds_{1,4} < ds_{3,4} < ds_{4,5} < ds_{1,2} < ds_{1,3} < ds_{2,5} < ds_{3,5}$									
ED ₀	1.00 < 2.24 ≤ 2.24 ≤ 2.24 < 2.83 < 3.16 < 4.24 < 5.00 < 5.39 < 5.83									

步骤 2. 顺序排列 SU 排序距离 这些数据来自 PCA, 对于这个初始布局, 前两个主分量坐标被用来计算所有 (i,j) 对的欧氏距离, 这些 SU 对间的欧氏距离按着由最低到最高的顺序排列[公式(21.4)].

顺序	1	2	3	4	5	6	7	8	9	10
对	$d_{2,3} < d_{2,4} \leq d_{1,5} \leq d_{1,4} < d_{4,5} < d_{3,4} < d_{1,2} < d_{1,3} < d_{2,5} < d_{3,5}$									
ED ₀	0.33 < 0.56 ≤ 0.57 ≤ 0.68 < 0.77 < 0.79 < 1.18 < 1.33 < 1.46 < 1.51									

注意这些 SU 对排序距离的顺序与群落数据中 SU 间的相异性的顺序(步骤 1)并不是十分匹配的, SU 对 (4,5) 和 (3,4) 以及 (2,5) 和 (1,3) 的顺序是相反的

步骤 3. 优化排列顺序. 为了获得正确的顺序, 在排序的布局中, SU₁ 和 SU₃ 的距离应该小于 SU₂ 和 SU₅ 间的距离, 我们可以通过变化在初始排列中的坐标, 即简单地移动使 SU_{s1} 和 3 接近, SU_{s2} 和 5 远离. 对于 SU_{s1,2,3,5} 在第一排序轴的位置, 如果我们加上或减去 0.05 个单位, 重新计算这个新的排序布局中 SU_s 间距离(表 21.2), 得到如下顺序的排列结果.

顺序	1	2	3	4	5	6	7	8	9	10
对	$d_{2,3} < d_{2,4} \leq d_{1,5} \leq d_{1,4} < d_{3,4} < d_{4,5} < d_{1,2} < d_{1,3} < d_{2,5} < d_{3,5}$									
ED ₀	0.26 < 0.60 ≤ 0.61 ≤ 0.64 < 0.74 < 0.83 < 1.18 < 1.36 < 1.43 < 1.52									

表 21.2 经过在原初 PC I 和 PC II 布局内反复移动 SU_s 以获得来自群落数据 SU 和 SU 距离顺序 [ds 公式(21.3)] 与 PCA 排序的 SU 和 SU 距离 [d, 公式(21.4)] 顺序的一致性, 进而得到来自 NMDS 排序的最终布局(排序)

SU	初始的*				最终的			
	I	II	I	II	I	II	I	II
(1)	-0.62		-0.28		-0.57		-0.28	
(2)	+0.55		-0.11		+0.60		-0.11	
(3)	+0.78		+0.12		+0.73		+0.12	
(4)	0.00		0.00		0.00		0.00	
(5)	-0.72		+0.28		-0.78		+0.28	

a 见表 19.3

注意这个 d_{ij} 的排列顺序现在与步骤 1 的 ds_{ij} 的排列顺序是完全一致的。

对于这个例子, 我们只是用我们的直觉来调整坐标, 而这种 NMDS 的排序算法需要用非参数回归计算应力 [SR, 公式(21.5)], 然后考虑到在每一维上对于每一个点要获得一个“负

梯度”，利用 SR 的部分导出值在这种布局内(调整的坐标)移动 SU_s，之后用人们所知道的阶梯函数(最速下降)在这种布局内移动这些点使之落在这个梯度上。正如早一些时候指出的那样，这些数学步骤不在本书范围之内。

步骤 4. 构造 NMDS 排序。作为最终的 NMDS 排序(图 21. 2b)，上面的排列(步骤 3)是可以接受的，注意这个排序和初始的 PCA 排序所期望的紧密相似性，在 21. 6 节中，我们将讨论初始的排列对于最终 NMDS 排序的影响。

表 21.3 使用程序 DPC. BAS 对巴拿马 6 个发生地(SU_s)在主分量 I 和 II 的格局进行无偏主分量分析的结果

变量	统计量			
	平均数	方差	标准差	标准误
X=PC I	-0.000	0.605	0.778	0.318
X ² =(PC I) ²	0.504	0.757	0.870	0.355
Y=PC II	-0.000	0.235	0.485	0.198
相关				
比较变量	相关系数	标准误	自由度	t 统计量
PC I, (PC I) ²	-0.922	0.193	4	-4.769
PC I, PC II	0.000	0.500	4	0.001
(PC I) ² , PC II	0.252	0.484	4	0.521

二次多项式回归方程

$$Y = -0.474 + (0.971X) + (0.941X^2)$$

多元测定系数 R²=0.426

在自由度(2,3)下 F 比率=1.115

21.4 例子:巴拿马蟑螂

使用由 PCA 对巴拿马蟑螂数据处理的结果(表 19. 4)，对这 6 个 SU_s 在 PC I 和 PC II (图 19. 5)的弓形进行了检验，运用 DPC 分析对其无偏化的可能性进行了分析。运行程序 DPC. BAS(软盘)，结果(表 21. 3)表明尽管这 6 个发生地 SU_s 在 PC I 和 PC II 上的格局有轻微的抛物线形状(图 21. 3a)，但不显著，发生地 3, 4, 5 的聚集强烈地影响抛物线的形状。

在 NMDS 排序中，这些蟑螂的数据再一次被利用，使用弦距离[公式(19. 4)]作为 SU 对间相似性的测定，来自 PCA 对于这些数据(表 19. 4)的最被排序布局(坐标)被用在程序 NMDS. BAS 中，经过一系列的反复迭代产生应力为一非常小的值后(即最终排序距离的排列顺序与群落数据 SU 和 SU 的相似性排列顺序非常匹配)，便产生了一个最终的旋转布局。注意这个 NMDS 排序与 PCA(图 19. 5)有实质上的不同，但与 COA 的排序有些相似的地方(图 20. 3)，这个差异可能是与我们利用欧氏距离获得 SU 相似性(ds)的排列顺序的这种事实有

关，也就是 NMDS 算法是在这个排列内通过移动 SUs 以试图达到对上述的排列顺序达到匹配。回忆一下，欧氏距离通过(SU)(列)隐含地对数据进行了标准化，而 PCA 是种的数积或相关也就是行标准化为基础的。另一方面，COA 使用的是双标准化(行和列)的数据，因此 COA 排序的排列方法与 NMDS 的更加相似些。Kenkel 和 Orloci (1986)指出关于 SU(列)标准化数据(如欧氏距离)的 NMDS 对于恢复沿着模拟的梯度由种“翻转”引起的各种程度的非线性性是最健全的方案。

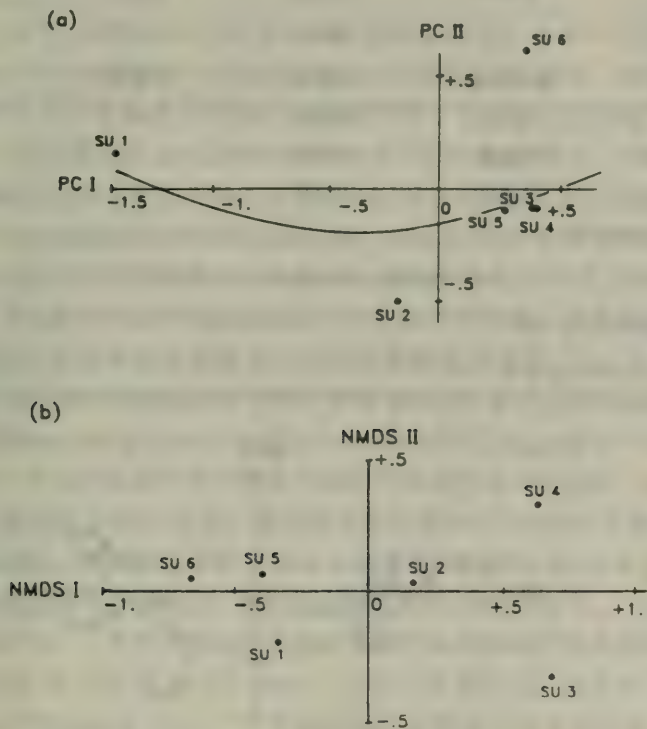


图 21.3 巴拿马 6 个发生地(SUs)的排序:(a)主分量分析表现出的 PC II 到 PC I 上的二次多项式回归的配合,和(b)SUs 在 NMDS 两个轴上 I 和 II 的非度量多维调节。

21.5 例子：威斯康星森林

使用 DPC. BAS 对威斯康星森林 10 个立地(表 19. 5,图 19. 6)PCA 产生的 SU 弓形的显著性进行了检验。表 21. 6 给出了这个程序的结果。多项式回归是显著的($P=0. 05$, $df=2, 7$), 如图 21. 4a 所示, 10 个 SUs 对抛物线作垂直投影得到 DPC 轴。这个 DPC 的结果与用无偏对应分析对这组同样的 10 个高地森林立地的分析结果是十分相似的(Gauch 1982, P. 158), 尽管我们的数据矩阵由 Gauch (1982, P. 122)所用的 14 种减少为 8 种。

同样森林数据的 NMDS 排序(在 NMDS. BAS 程序中再次利用欧氏距离和 PCA 初始的排列)产生了一个没有旋转的 SUs 格局, 很明显, 同 PCA(图 19. 6)或 COA(图 20. 4)的 SU 格

表 21.4 利用 DPC, BAS 程序对于 10 个森林立地的主分量 I 和 II 上的格局进行
无偏主分量分析的结果

变量	统计量			
	平均数	方差	标准差	标准误
$X = PC\ I$	0.000	0.497	0.705	0.223
$X^2 = (PC\ I)^2$	0.447	0.054	0.232	0.073
$Y = PC\ II$	0.000	0.153	0.391	0.124
比较变量	相关			
	相关系数	标准误	自由度	t 统计量
$PC\ I, (PC\ I)^2$	-0.111	0.351	8	-0.315
$PC\ I, PC\ II$	0.000	0.354	8	0.001
$(PC\ I)^2, PC\ II$	0.833	0.196	8	4.259

二次多项式回归方程

$$Y = -0.635 + (0.051X) + (1.419X^2)$$

多元测定系数 $R^2 = 0.426$

在自由度 (2,3) 下 F 比率 = 1.115

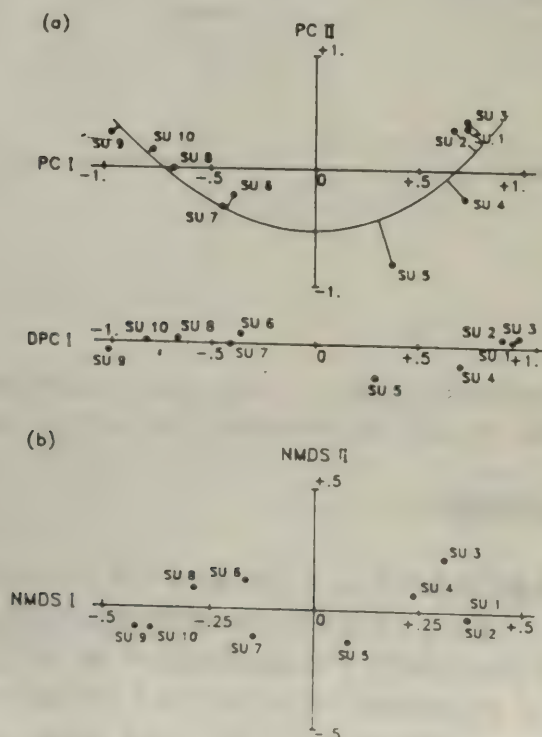


图 21.4 10 个高地森林立地(SUs)排序:(a)关于 PC I 和 PC II 的多项式排序(DPC)和通过把这些 SUs 向抛物线上作垂直投影而形成的无偏主分量轴(DPC I)。和(b)非度量多维调节显示的 SUs 在 NMDS 轴 I 和 II 上的位置

局相比它很少有弓形,换句话说,就是在获得它们排列顺序距离与 SU 和 SU 间弦距离尽可能匹配的排序布局下,NMDS 算法重新放置这 10 个 SUs。这种重新放置的 SUs 到一个排列中稳定后,它的第一排序轴(NMDS I)通过 SU 1 和 SU 2 到 SU 9 和 SU 10 的一个森林梯度是非常接近于一条直线的(见 Gauch 1982, Peet 和 Loucks 1977)。

21.6 非线性排序的附加讨论

正如 17 章讨论的那样,排序的目的就是在种的多维空间内放置 SUs 以使它的空间轴代替潜在的环境梯度,它将有希望导致一些有意义的生态解释(24 章)产生。当种的关系沿着这些梯度是非线性(如钟形曲线的重叠)的时候,问题就产生了,这种人为的就不能代表真正潜在的梯度(如图 21.1),而且梯度可以以曲线形式的弓形(二维)或螺旋形(三维)结束。这一章里,我们提供了可以展开这种弓形的无偏主分量(DPC)的过程,之后我们又提供了非度量多维调节,它是一种对弓形或螺旋形作顺序排列的过程。很明显弓形/螺旋形问题又真正回到了正确地估测生态距离的相似性方法的失误上,当 SUs 通常仅有几个种或没有时(即出现在宽的梯度对立的两端),和(或)种的关系是非线性时(也就是当种沿着一较宽的梯度分布时),这种失误尤其尖锐,排序方法的早期发展解释这种失败有些地方是成功的,但有限(见 Gauch 1973, Gauch 等 1974, Ihm 和 Groenewond 1975)。不管怎样,在排序中准确地使用相似性方法的研究工作还在继续(Bradfield 和 Kenkel 1987, Faith 等 1987),Feuster 和 Orloci 1983)。我们建议学生们经常注意一下包括以模糊数学理论为基础的非线性排序方法的一些新发展。

在这章里我们选择提供了 DPC 过程,用以使经常发现 SUs 在由 PC I 和 PC II 形成的坐标内表现出的弓形无偏化,我们也注意到在其它分量组合(如 PC I 和 PC III, PC II 和 PC III)上出现的弓形也可以通过这个步骤测定,正如 Phillips(1976)表示的那样。

我们还提供了 NMDS,它在处理种沿着陡而长梯度的钟形曲线产生的非线性方面正在被认为是一种健全的方案(Kenkel 和 Orloci 1986)。在圣卡他利那山,亚利桑那州(Niering 和 Lowe 1984, Whittaker 和 Niering 1964),种沿着陡峭的湿度梯度分布的这种曲线是明显的。发展的一些其它排序方法包括高斯排序法(Gaussian ordination method [Gauch 等 1976]),在这个方法中,其 SUs 是沿着第一排序轴放置的,这样就使得沿着这个轴每个种的格局最好地配合高斯曲线。这个过程以种的 SUs 近似于钟形曲线的初始放置开始,之后 SUs 的位置被反复地调整以获得所有种对高斯曲线的最佳配合。

Ihm 和 Groenewond (1975)和 Johnson 和 Goodall (1979)描述了以高斯模型为基础的类似过程。作为对这个模型的一个修改,Orloci (1980)提出了一种排序算法,它允许用除了高斯以外的其它种-多度模型。基本的高斯排序方法在处理具有单一优势梯度的数据时是有效的,但在试图扩大这个模型到多个轴时,结果通常不令人满意(Gauch 1982)。

Noy-Mein (1974)描述了一种与 NMDS 有关的非线性排序方法,他称这个过程为连续体分析(Continuity analysis),因为 SUs 按着 Noy-Mein 所说的“部分相似性(local similarities)”最优化方式的这种连续系列中反复地排列,这种过程是由心里测验过程中人们所知道的参数图距导出的(Shepard 和 Carroll 1966)。但连续体分析似乎只在处理具有中等程度的非线性种-多度数据时表现较好(Gauch 1982),这里种-多度曲线只是钟形曲线的部分

片断(如图 21. 1c)。

Bradfield 和 Kenkel (1987)介绍了另一种与非线性排序有关的方案。他们的方法称为最短路径调节(Flexible shortest path adjustment),包括在 SU 和 SU 距离矩阵中更换一定的值。这些更换称“最短途径”,并代表在种类组成上只有一点或少数重叠的 SUs 间重新计算的距离,这种技术的优点使用者有灵活性决定一些条件,计算最短路径以获得最有效的排序。

NMDS 算法从 SUs 在坐标系统内的最初布局开始,之后在这个布局内移动 SUs,以使 SU 和 SU 间距离的排列顺序尽可能与数据矩阵中的 SU 和 SU 间的相似性匹配。一些比较研究表明最初排列方法的选择并不十分重要的,由于这种算法通常汇集于最优化的解法(Fasham 1977),但我们的经验表明,局部优化很可能与小的数据组,甚至有可能与大的数据组有关。因此,我们推荐使用一些不同的初始排列方法,包括随机排列(Kenkel 和 Orloci 1986)。Fasham (1977)推荐至少使用 2 个最初的排列方法;随机格局和另一个以 COA 排序为基础的排列方法。其目的是在 NMDS 中使用不同的最初排列方法获得相似的而不是统一的最终的 SU 排列。(回忆一下这种 NMDS 方法仅以优化排列顺序为基础,而不是实际的可度量距离)。NMDS 的不同计算机程序产生的最终排列,它们在调节和旋转方面不同。

NMDS 算法也对在这个排列方法中的轴数或维数优化,就是两维 NMDS 的排序 SU 的格局将不同于三维 NMDS 排序前两维 SU 的格局。因此,重要的排序轴数(即明显的潜在梯度的数目)必须知道或预先指定,当然收集群落数据的生态学家是很有可能能够指定这个的。如果不能确定,可使用较高和较低的维数重复 NMDS,最后选择一个能提供最好生态解释的解法。

当然,本书不可能估价和比较所有不同的排序方法,学生们可以参见 Gauch (1982) Minchin (1987), Orloci (1978), 和 Pielou (1984)的杰出总结。Pielou (1984)指出不同的排序方法“在恰当的条件下各有其优点”,这个阐述强调一个排序方法的选择(或不同方法)依赖于手头上数据的特征和心里的目标。

21.7 总结和推荐

1. 沿着长或陡的梯度取样获得的群落数据典型的包括种间非线性重叠的钟形种-多度曲线,这类数据导致在排序过程中 SU 格局强烈地呈弓形、圆形和螺旋形。

2. 当主分量分析(PCA 19 章)表明 SUs 在这个排序中呈弓形格局时,这个弓形可以用多项式回归而被无偏化或展开,这种过程称多项式回归或无偏主分量(DPC)分析,我们建议这个过程逻辑上接着 PCA 进行。

3. 当对应分析(COA 20 章)表明由于种间的非线性关系导致一弓形格局出现时,我们建议使用无偏对应分析(DCA)作为拉直这种弓形的方法(Hill 和 Gauch 1980),尽管 DCA 的算法十分复杂,并不在本书范围内,但计算程序还是可以利用的(如 Cornell 生态程序;见 Gauch 1982, P. 256)。

4. 在一个排序内当 SU 格局很可能是较强的非线性时(如圆形或螺旋形),我们建议用非度量多维调节的排序方法(NMDS)(21. 2. 2 节),比较研究支持 NMDS 过程使用 Q-方式相似性方法[ds, 公式(21. 3)],它通过 SU 对数据标准化(如弦距离, Orloci 和 Kenkel 1986, 或

百分相异性, Minchin 1987)。

5. 目前的生态相似性方法(如距离或相似性, 14 章)不能恰当地测定位于梯度端点 SUs 有真正分离(21.6 节), 方法的研究还在继续, 以改进排序的工作(如 Bradfield 和 Kenkel 1987)。

6. 在分析之前, 给定 NMDS 排序的维数必须被指定, 因为 NMDS 算法依赖于固定的维数(21.6 节), 对于重要的潜在梯度的数目如不能确定, 我们建议重复 NMDS, 最终维数的选择要以生态解释容易的一个为基础。

第七部分 群落的解释

第二十二章 背景

在前九章中,我们介绍了各种群落分类和排序的方法。总的来讲,分类的目的是缩小 S 种 N 个取样单位(SU)数据矩阵,把它分成 $g < N$ 个组,此处各组内 SU 之间的相似性大于各组间 SU 的相似性,而排序的目的是把 SU 数据矩阵转化成一个缩小的 $K < S$ 的排序系统(Green 1980)。换句话说,分类的目的是把 SU 聚类或划分组成同质性组,而排序的目的是把 SU 在一个轴系统内排列,以反映相似性格局。实际存在的、可能对最后分类或排序格局有影响的环境因子,并未直接参与我们的分析,虽然对这些环境因子的考虑会影响我们对特定方法的选择。在以下两章中,我们将探索环境因子对群落分类(第二十三章)和排序(第二十四章)格局发生影响的过程。

对于一个分类结果给出的 g 个组或群体的环境解释,逻辑上必需使用一些统计方法以检验在每一组内测量的环境因子(变量)之间的差异。例如,这些方法可能包括非参数秩(Non-parametric rank)和符号检验(sign test)或参数单径方差分析(parametric one-way analysis of variance)和平均值对比检验(mean comparison test)。经常处理许多同时因子的生态学家,可以利用多变量统计方法。尤其是为了检测群落差异(第二十三章),而同时检测多个因子时,判别分析是一组很有用途的方法。我们把判别分析(discriminant analysis)分成两类:简单的和复合的。前者用于组对之间比较,后者则用于同时比较所有 g 个组。在第二十三章中,我们将介绍简单判别分析(SDA)的过程。

由一个排序结果,给定了 N 个 SU 沿 $K < S$ 个轴(梯度)的排列,这个排序的环境解释,在逻辑上必需利用把这些轴和在每个 SU 中测定的环境变量相联系的方法。这些方法包括利用简单和复合线性回归,把排序轴和环境因子联系起来(第二十四章)。

22.1 矩阵展示

群落解释可以被看作是一个两步的过程。首先,用 SU 中的种丰富度或种有一无数据进行群落分类或排序。下一步是利用每个 SU 中测得的环境因子(例如,土壤水分和 PH 值)数据,对群落格局加以解释(图 22.1)。注意我们现在用的是完整的数据矩阵,即包括种类数据子集和环境因子数据子集的所有行和列。

22.2 选择的文献

简单和复合判别分析被广泛用于许多群落——溪流、荒漠和森林(表 22.1)的环境解释。在生态学中,最早应用判别分析的研究之一是在一个山毛榉林内进行的,此处用土壤数据对群落作了环境解释(Norris 和 Barkham 1970)。利用复合线性回归解释排序结果已被生态学家

使用过，但大部分都是在地生物群落中。

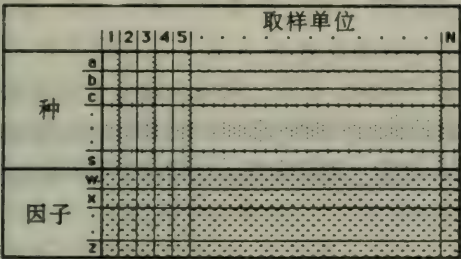


图 22.1 阴影部分表明了群落解释研究的生态学数据矩阵的形式。我们感兴趣的是用于分类和排序中的种-SU 数据，以及 SU 中环境因子数据的利用

表 22.1 利用简单判别分析(SDA)，复合判别分析(MDA)和复合线性回归(MLR)进行群落解释研究，并给出与示例的有关文献

地点	群落	方法	参考文献
蒙大拿	蓝松鸡	SDA	Martinka 1972
弗吉尼亚	啄木鸟	SDA	Conner 和 Adkisson 1976
阿拉斯加	萨蒙	SDA	Swanston 等 1977
英国	山毛榉林	MDA	Norris 和 Barkham 1970
纽约	水生生物	MDA	Walker 等 1979
西德	植物	MDA	Wiegleb 1981
亚利桑那	荒漠鸟类	MDA	Rice 等 1983
波·德尔塔(意)	栎林	MDA	Gerdol 等 1985
加利福尼亚	红树林	MLR	Waring 和 Major 1964
怀俄明	高山苔原	MLR	Scott 和 Billings 1964
新不伦瑞克(加)	森林	MLR	Forsythe 和 Loucks 1972
雅里(西藏)	平原—山地	MLR	Chang 和 Gauch 1986

第二十三章 分类的解释

在这一章里我们介绍简单判别分析(SDA)，这是一种检验由分类划分出的群体中非生物特征显著差异的方法。通过计算多变量距离(D^2)统计量和 F 比率，SDA 使我们可以对群落间的显著差异进行检验。另外，我们可以计算一判别函数，这对估计每一变量对群落中环境判别的相对贡献大小是有用的。

23.1 基本方法

Gauch (1982)列举了生态学中分类的三个目的：(1)总结大而复杂的数据集；(2)有助于群落变异格局的环境解释；(3)改进群落结构的模型。

在第四部分(第十五、十六章)中介绍了把 SUs 集合进行分组或聚类的关联分析和聚类分析。我们知道分类过程可以被看作是把 S 行(种)， N 列(SUs)的数据矩阵转化成 $g < N$ 个组，此处每组内 SUs 之间的相似性大于每组之间 SUs 间的相似性(Green 1980)。群落生态学的一般目的就是取自不同环境中的样本中区分出 g 个同质群体。给出一个 $g < N$ 结构的数据集，我们现在可以开始有关生态学解释的问题的探讨。例如，是否由于不同的种反应，存在引起 SUs 分成 g 组的环境因子？如果存在，这在生态学上意味着什么？

表 23.1 参量和非参量统计检验在比较组(群体)内定性和定量环境因子之间的差异中是有用的

组数	检验类型	检验名称	评价
2	参数	t-统计	广泛应用于平均差异的单变量检验
	参数	Hotelling T^2	相当于 t-统计的,用于多变量的检验
	参数	简单判别分析	F-检验和 Hotelling 的 T 有联系
	非参数	符号检验	可用于成对的定性数据
	非参数	Wilcoxon	非成对定量数据,检验秩的差异
≥ 3	参数	有平均值比较的	广泛用于所有差异和组
		单径方差分析	平均对比的单变量检验
	参数	多变量判别分析	平均向量差异和多变量距离的多变量检验

为了检验从 g 组内的每个 SU 中测定的一系列环境因子在各群体之间的显著差异，必需选择一个适当的统计检验方法。在表 23.1 中，给出了一些可能的检验。显然这个表并不完全，但已列出了一些重要的可以应用于生态学的检验方法。

参数检验有大量与被比较总体参量的统计评价相联系的假设。例如，假定 SUs 是取自一个无限总体的随机取样，因此和每个被测定的环境变量(例：土壤结构或含水量)相联系的误

差项是独立的正态分布变量。如果不坚持这个假设,那么在某些情况下,进行数据变换去迎合这个假设是可能的。另一个选择是非参数(自由分布)检验。采用非参数检验的优点是摆脱了这样的参数假设,而且计算比较容易。然而,非参数检验不如参数检验那样有效(例如,缺乏说服力),然而,如果坚持这个假设、非参数检验则不如参数检验那样有效(例如,缺乏说服力)。

在这一章中,我们介绍了简单判别分析(SDA),它是检测 g 组(群体)内环境差异的有用工具。同学们可能最熟悉检测组间差异的单变量的 t -统计检验和方差分析(AOV)检验,但可能对 SDA 不太熟悉。该过程十分有效,并且高度可变,在这个过程中完整的 SDA 产生的函数可用于(1)检测多个同时环境因子(多变量 AOV)在 g 组之间是否存在显著差异;(2)检测组之间的多变量距离;(3)检测 g 组中每一组内每一环境因子的相对重要性;(4)基于环境特征,给出一个新 SU 归入已存在的 g 个组之一中的最大可能性分类。从历史上看,SDA 在生物学数据上的应用可以追溯到 Fisher (1936)的判别函数和 Mahalanobis (1936)的 D^2 多变量距离的同时发展。虽然后者可由前者计算得到,但判别函数和 Mahalanobis D^2 距离的应用方面可以很不同(例如, Fisher 检测鸢尾属(Iris)的植物分类,而 Mahalanobis 研究人类学的不相似性)。因为它最初来源于 Fisher 和 Mahalanobis,因此 SDA 和多元线性回归有关。一个 SDA 的图表纵览在图 23.1 中给出。从图上可以看到两组 SUs(标有 1 和 2)被标在两个环境变量(X_1 和 X_2 ,水平轴和垂直轴)空间内。在此我们假设,基于种类丰富度的聚类分析,划分出了两 SU 组,每个组由 7 个 SU 组成。从图 23.1 中可以看到,即不是 X_1 也不是 X_2 ,可以独自把两组区分开,即当把 14 个 SUs 分别投影到 X_1 和 X_2 轴上时,1 组和 2 组的 SU 在轴上有重叠。然而,判别函数——(Z)- X_1 和 X_2 两者的线性结合(一个和 X_1 和 X_2 成一定角度的 Z 轴)才能把两个 SU 组区分开。

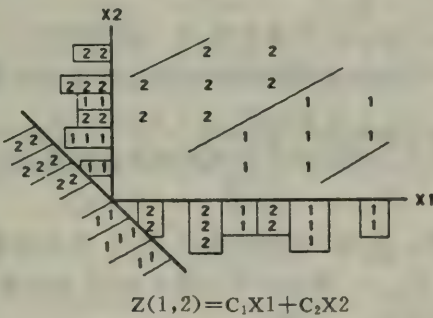


图 23.1 简单判别分析的图示。在环境因子 X_1 和 X_2 的空间内不重叠的两个 SU 组 1 和 2。
仅基于 X_1 和 X_2 区分出这两组是困难的。用于区分两组的新判别函数(Z)是 X_1 和 X_2 的线性结合。Modified 摘自 Legondre 和 Legendre (1983)

23.2 步骤

给出一 N 个 SUs, S 个种的数据集,假设用分类方法区分出 $g < N$ 个组。在每一个 SU 中测定了许多环境变量 X_i , ($i=1, \dots, P$)。为了简单起见,下面叙述的 SDA 过程限制在 $g=2$ 组或群体(1 和 I), $P=2$ 个环境变量(X_1 和 X_2)的情况。在 I 和 I 组内的 SU 个数分别是 N_1 和

N_1 。在这个过程中，不论组数(g)和环境因子数(P)是多少都可以使用。但是象我们上面提到的，在SDA中每一组是以一对一的形式和其它所有组进行比较的，因此，以下叙述的步骤，对于每一组对就必须重复一次。

第一步：计算因子的平均值、方差和协方差。在组 I 和 II 中，计算两个环境变量(X_1 和 X_2)的平均值(\bar{X}_1 和 \bar{X}_2)、方差(S_1^2 和 S_2^2)和协方差($S_{1,2}$) (见 Sokal 和 Rohlf 1981)。

第二步：计算联合方差—协方差矩阵。联合矩阵的方差和协方差项的计算包括累加两组的平方和及交积两项，再除以自由度($N_1 + N_2 - 2$)。计算方法在 Pielou (1984, 表 3, 2) 和大多数统计课本中都有说明。联合矩阵如下：

$$S^* = \begin{bmatrix} S_1^{*2} & S_{1,2}^* \\ S_{2,1}^* & S_2^{*2} \end{bmatrix} \quad (23.1)$$

此处星号(*)表示组 I 和 II 的联合值， S_1^{*2} 和 S_2^{*2} 是环境因子 X_1 和 X_2 的方差， $S_{1,2}^* (= S_{2,1}^*)$ 是两个变量之间的协方差。

第三步：计算判别函数的系数(C_1 和 C_2 ；图 23.1)。首先，我们建立一个线性方程系统：

$$C_1 S_1^{*2} + C_2 S_{1,2}^* = d_1 \quad (23.2a)$$

和

$$C_1 S_{2,1}^* + C_2 S_2^{*2} = d_2 \quad (23.2b)$$

此处 $d_1 = \bar{X}_{1,1} - \bar{X}_{1,2}$ 和 $d_2 = \bar{X}_{2,1} - \bar{X}_{2,2}$ ， d_1 和 d_2 是组 I 和 II 的平均向量差值。我们这个两组和两个变量的简单示例的系数 C_1 和 C_2 可以用下式求得：

$$C_1 = \frac{S_2^{*2} d_1 - S_{1,2}^* d_2}{V} \quad (23.3a)$$

$$C_2 = \frac{S_{1,2}^* d_2 - S_{2,1}^* d_1}{V} \quad (23.3b)$$

此处： $V = [(S_1^{*2})(S_2^{*2}) - (S_{1,2}^*)(S_{2,1}^*)]$ ，它是方程 23.1 的判别式。

第四步：两个组，I 和 II 之间的总判别函数(2)可以写成如下形式

$$Z_{1,1} = (C_1)(X_1) + (C_2)(X_2) \quad (23.4)$$

第五步：计算 Mahalanobis D^2 。组 I 和 II 之间的多变量距离(D^2)是

$$D_{1,1}^2 = (C_1)(d_1) + (C_2)(d_2) \quad (23.5)$$

第六步：检验组矩心差异。在两个环境变量 X_1 和 X_2 的多变量空间内，组 I 和 II 的矩心分别是 $(\bar{X}_{1,1}, \bar{X}_{2,1})$ 和 $(\bar{X}_{1,2}, \bar{X}_{2,2})$ 。第五步中计算出的 D^2 值可被用于检验两个组矩心无差异的假设。F 统计可由下式计算(Sneath 和 Sokal 1973)：

$$F_{1,1} = \frac{D_{1,1}^2 [N_1 N_2] [N_1 + N_2 - 3]}{2 [N_1 + N_2] [N_1 + N_2 - 2]} \quad (23.6)$$

这个 F—统计值和自由度为 $[2, N_1 + N_2 - 3]$ 的临界 F 值表进行比较。

第七步：决定 X_1 和 X_2 对 D^2 的相对百分比贡献。 X_1 和 X_2 对组 I 和 II 之间总多变量距离的相对百分比贡献，分别用 RC_1 和 RC_2 表示，二者可用下式计算：

$$RC_1 = \frac{C_1 d_1}{D^2} \quad (23.7a)$$

$$RC_2 = \frac{C_2 d_2}{D^2} \quad (23.7b)$$

第八步：用图表示判别函数(Z)。基于判别函数 $Z_{1,1}$ 建立一个轴，并在该轴上标出组矩心的参考得分。这些组矩心值可以通过把环境变量 X_1 和 X_2 的平均值代入Z得到：

组 I : $Z_1 = (C_1)(\bar{X}_{1,1}) + (C_2)(\bar{X}_{2,1})$ (23.8a)

组 I : $Z_1 = (C_1)(\bar{X}_{1,1}) + (C_2)(\bar{X}_{2,1})$ (23.8b)

当然，每一SU在轴上的位置也可以由每个SU中的 X_1 和 X_2 代入 $Z_{1,1}$ 中得到。

23.3 举例：计算

对表 11.3 中的人为数据(五个SU,三个种)进行了扩展，又包括了两个环境变量数据 X_1 和 X_2 (表 23.2)。基于聚类分析(见图 16.2)，五个取样单位被分成了 $g=2$ 个组(I和II)。这些数据的SDA计算在下面加以说明。

表 23.2 五个SU中的两个环境因子值和三个种丰富度的生态学数据矩阵。由聚类分析区分出两个组或群体

		组				
变量	SU _s :	I		II		
		(1)	(5)	(2)	(3)	(4)
种	(1)	2	0	5	5	3
	(2)	0	1	3	4	2
	(3)	2	2	0	1	0
环境因子	X_1	4.2	3.8	1.6	2.4	2.0
	X_2	6.1	5.9	7.7	8.3	8.0

表 23.3 组 I 和 II 中环境因子 X_1 和 X_2 的平均值、方差和协方差

		组 I					
		SU _s		平均值 \bar{X}	方差 S ²	协方差 S _{1,2}	
		(1)	(5)				
环境因子	X ₁	4.2	3.8	4.0	0.08	0.04	
	X ₂	6.1	5.9	6.0	0.02		
		组 II					
		SU _s			平均 \bar{X}	方差 S ²	协方差 S _{1,2}
		(2)	(3)	(4)			
环境因子	X ₁	1.6	2.4	2.0	2.0	0.16	0.12
	X ₂	7.7	8.3	8.0	8.0	0.09	

第一步：表 23.3 中给出了平均值、方差和协方差。

第二步：联合方差-协方差矩阵(方程 23.1)是

$$S^* = \begin{bmatrix} 0.133 & 0.093 \\ 0.093 & 0.067 \end{bmatrix}$$

第三步：计算判别函数的系数。根据方程 23.2，我们建立了线性方程系统。

$$C_1(0.133) + C_2(0.093) = 4 - 2 = +2.0$$

$$C_1(0.093) + C_2(0.067) = 6 - 8 = -2.0$$

然后，由方程 23.3

$$C_1 = [0.067 \times 0.2 - 0.093 \times (-2.0)] \div V = 1800$$

$$C_2 = [0.133 \times (-2.0) - 0.093 \times 2.0] \div V = -2550$$

此处 $V = 0.133 \times 0.067 - 0.093 \times 0.093 = 0.000178$

第四步：写出判别函数，Z(23.4 方程)

$$Z = (1800)(X_1) + (-2550)(X_2)$$

第五步：Mahalanobis D^2 统计[23.5 方程]是

$$D^2 = 1800 \times 2.0 + (-2550) \times (-2.0) = 8700$$

第六步：检验组矩心的差异。F 检验统计是(23.6 方程)：

$$F = [8700 \times 3 \times 2 \times (3 + 2 - 3)] \div [2 \times (3 + 2) \times (3 + 2 - 3)] = 3480$$

只基于 $F=3480$ ，自由度 $df=(2,2)$ ，获得 $D^2=8700$ 的概率小于 0.001。我们放弃组 I 和 II 矩心之间无差异的零假设。

第七步：决定 X_1 和 X_2 对 D^2 (23.7 部分)的相对百分比贡献：

$$RC_1 = 1800 \times 2 \div 8700 = 0.41 \text{ 或 } 41\%$$

$$RC_2 = -2550 \times 2 \div 8700 = 0.59 \text{ 或 } 59\%$$

因此，我们可以得到结论：在区分组 I 和 II 时，环境因子 X_2 比 X_1 的重要性大 20% 左右。

第八步：图示判别函数(Z)。组 I 和 II 的参考得分是[23.8 方程]：

$$Z(I) = -8100 \text{ 和 } Z(II) = -16800$$

把 5 个 SU 的每一个 SU 中的 X_1 和 X_2 值代入(23.4 方程)，各 SU 定位于 Z 上

$$Z(1) = -7995, Z(5) = -8205, Z(2) = -16755$$

$$Z(3) = -16845, Z(4) = -16800$$

从平均值和沿判别函数 Z 上组 I 和 II 中的 SU 的参考得分图可以看出，基于多变量距离的组 I 和 II 的分离。

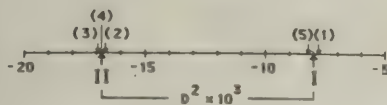


图 23.2 组 I 和 II 之间的判别轴，显示了组平均值的位置和各 SU 位置。同时显示了两组之间的 D^2 距离

为了寻找可能的环境解释，我们检测了由分类产生的组 I 和 II。两个组内的环境变量 X_1

和 X_2 有明显差异，并可用 X_1 和 X_2 去区分两个组，其中 X_2 的重要性大于 X_1 的重要性。

表 23.4 巴拿马六个地点的五种蟑螂和两个环境因子的生态学群落数据矩阵(来自 Wolda 等 1983)

种	数值	地点					
		BCI	LC	FORT	BOQ	MIR	CORG
		(1)	(2)	(3)	(4)	(5)	(6)
Ceuthobiella spp.	(1)	0	0	0	0	0	1
Compsodes cucullatus	(2)	14	38	1	1	4	0
Compsodes delicatulus	(3)	28	4	1	0	1	0
Buboblatta armata	(4)	7	0	0	0	0	0
Latindia dohrniana	(5)	68	29	0	0	11	24
环境因子							
海拔(km)	(1)	0.120	0.140	1.050	1.350	0.005	0.100
降水(m)	(2)	2.5	1.5	5.0	2.5	3.0	2.5

23.4 例子：巴拿马蟑螂

巴拿马六个地点蟑螂的丰富度的数据集(表 11. 4a)可以被扩展成包含两个环境因子—海拔和降水量信息的数据集，由 Wolda 等(1983)提供。这些数据在表 23. 4 中给出。

利用聚类分析(16. 3 部分)进行的蟑螂数据的分类表明：六个地点可以被分成相对同质的两组：组 I 由地点(1)、(5)和(6)组成，组 II 由地点(2)，(3)和(4)组成。我们将用这两个 SU 组说明应用 BASIC 程序 SDA. BAS 计算 SDA 的过程。然而，可以注意到关联分析也区分出两组，但是把地点(6)和其它五个地点分离，虽然我们的组的选择是人为的，在分组时仍考虑到排序结果。极点排序中地点(3)和(4)位置很近，而和其它四个地点远离，PCA 排序中地点(3)、(4)和(5)被放得很近；在 PCA 排序中地点(3)、(4)和(5)被放在一起，而和地点(1)、(2)及(6)远离。

这些数据 SDA 的结果表明，基于海拔和降水量的两个蟑螂组之间区别不明显(表 23. 5)。事实上，这个结果说明了 SDA 的一个有趣的特性，包括第二个因子(降水量)实际上减弱了分析，因为它对多变量距离的相对百分比贡献是负值(−6. 2%)。因此，把一个弱判别变量包括在 SDA 中实际上可以减弱分析；在多元线性回归(Sokal 和 Rohlf 1981)中不存在这种情况，该方法中加入的变量总是增大分析的力量(虽然，常常没有多大意义)。

两个蟑螂组(I 和 II)在线性判别函数上的位置已在图 23. 3 中标明，包括六个地点中每一个地点的位置。组 II 中各地点(2、3 和 4)参考得分的范围进一步证明，把海拔和降水量作为因子的判别效果是不好的。利用各组参考得分之间的中点作为最小可能错误分类的划分线(Sneath 和 Sokal 1973)。地点(2)实际被错误地归入组 I。

表 23.5 利用海拔(X_1)和降水量(X_2)解释两个蟑螂地点组之间的差异, SDA.BAS 程序的判别分析结果[各组内环境因子的平均值(Mean)、方差(Var)、标准差(SD)、标准误(SE)和协方差(Covar)统计]。

组	因子	Mean	Var	SD	SE	Covar
I	X_1	0.075	0.004	0.061	0.035	-0.018
	X_2	2.667	0.083	0.289	0.167	
II	X_1	0.847	0.397	0.630	0.364	0.607
	X_2	3.000	3.250	1.803	1.041	

判别系数 $c_1 = -4.809$ 和 $c_2 = 0.651$

Mahalanobis 距离 = 3.494

df = (2,3) 的 F 统计量

每个环境因子与贡献

环境因子	相对贡献
X_1	106.2
X_2	-6.2

判别轴上的组矩心

$$Z_1 = 1.38 \text{ 和 } Z_1 = -2.12$$

SU 在判别轴上的得分

组	I			II		
	(1)	(5)	(6)	(2)	(3)	(4)
SU 得分	1.05	1.93	1.15	0.30	-1.79	-4.86

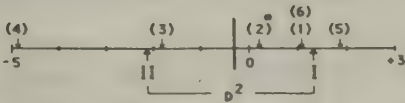


图 23.3 巴拿马六个地点蟑螂的简单判别分析表明组 I 和 II 的参考得分, 和判别函数上六个地点的位置。此处也显示了 Mahalanobis D^2 距离。地点(2)被错误分类。

23.5 例子：威斯康星森林

我们扩展了 10 个高地森林地点(表 11.6a)8 个树种的数据矩阵, 使它包括了 Peet 和 Loucks(1977)的地点因子(表 23.6)的信息。他们提供了土壤结构信息(A1 层中五级含沙百分比), 他们把该信息视为表示温度—养分状况的指数。他们还提供了位置动态学指数, 该指数

被视为种周转率的指数；这个指数和地点特性有关。

表 23.6 威斯康星南部 10 个高地森林地点两个地点因子的生态学数据矩阵。这些因子和相应的树种数据的详细情况见 Peet 和 Loucks(1977)和 Gavch (1982)

地点因子	数目	森林地点(SU _s)									
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
土壤质地	(1)	4	5	3	2	1	1	2	1	1	1
立地动态	(2)	1	2	1	2	1	3	3	5	4	5

聚类分析(图 16. 4)的结果被用于把 10 个 SU 分入两组内:组 I 包括 1-4 地点,组 II 包括地点 5-10。结果表明,利用土壤结构和位置动力学因子(表 23. 7),表明两组之间存在显著的多变量距离($D^2=9.5$)。两组矩心相等($F=10.0$)的概率小于 1%。土壤结构在区分两个组时的重要性是位置动力学指数重要性的三倍。

表 23.7 威斯康星南部两个高地森林 SU 的群体或组之间的 SDA, 判别采用的两个因子: X_1 =土壤结构(五级含沙量百分比), 和 X_2 =种周转率(见 Peet 和 Loucks 1977)表示的位置动力学指数。各組内环境因子平均值(Mean)、方差(Var.)标准差(SD)标准误(SE)和协方差(Covar)

组	因子	Mean	Var	SD	SE	Covar
I	X_1	3.50	1.67	1.29	0.645	<0.001
	X_2	1.50	0.33	0.58	0.289	
II	X_1	1.17	0.17	0.41	0.167	-0.100
	X_2	3.50	2.30	1.52	0.619	

判别系数 $c_1=3.10$ and $c_2=-1.16$

Mahalanobis 距离=9.547

df=(2,7)时 F 统计量为 10.0

贡献

环境因子	相对贡献%
X_1	75.8
X_2	24.2

判别轴上的组矩心

$Z_I=9.1$ 和 $Z_{II}=-0.4$

SU 在判别轴上的得分

组:	I					II				
SU:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
得分:	11.2	13.2	8.1	3.9	1.9	2.7	-0.4	-2.7	-1.5	-2.7

组参考得分数和 10 个地点(SU)在判别函数(图 23.4)上的位置说明了两个组的分离。注意地点并未紧密地聚集于组 I 参考得分的附近,事实上,基于最大可能临界值,地点(4)被错误地分入组 I。

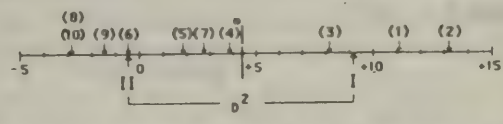


图 23.4 10 个威斯康星森林地点的简单判别分析表明组 I 和 II 及 10 个地点在判别函数上的分布。
Mahalanobis D^2 距离也标明了。* 地点(4)被错误地分类了。

23.6 分类解释的附加讨论

SDA 中每一组或群体和其它组或群体的比较方式是一对一对的,即决定每一个相互对比的组对的判别函数, D^2 统计是计算每对之间的多变量距离, F 比率用于检验这个距离的统计学显著性。然而,我们的兴趣集中在比较三个或更多同时存在的群体(例如, Gerdol 1985, Walker 等 1979, wegleb 1981)。在这种情况下,判别函数,多变量距离和 F 比率包括了所有组之间距离之和。这种方法一般称作——复合判别分析(MDA)或经典变量分析。

MDA 涉及到联合方差—协方差矩阵特征向量和向量的计算,在本章内没有作说明。然而,同学们会发现,在获得了此处介绍的简单 SDA 过程的感性认识后,转向三组或更多组的同时比较(MDA),就相对比较容易了。SDA(两组)可以得到一个判别函数轴,MDA(三或更多组)可以得到两个或更多判别函数(经典的)轴。我们推荐同学们参阅一下 Gittens (1985), Legendre 和 Legendre (1983), Orloci 和 Kenekl (1985), Pielou (1984) 和 Williams (1983) 的 MDA 处理。

在第二十二章中,我们说明了分类可以考虑两步过程:(1)基于种丰富度把 N 个 SU 缩小为 $g < N$ 个组和(2)检验 g 组内环境因子的统计学差异。这看起来是一个合乎逻辑的过程,即首先用生物区系划分群体,然后检测非生物因子的差异,这有助于解释生物反应。第二步便是本章重点介绍过的,许多单变量和多变量统计检验在决定 g 个组时很适用于检测种的重要性。用类推的方法,所有提出的有关非生物因子的问题可以用解决 S 种(生物)的方法来解决。例如, g 组的种矩心是否有显著差异?在决定各组之间 D^2 时每个种的相对贡献是多少?Walker 等(1979)进行了这样一项研究,利用水底生物种丰富度判别基层分类的五个组。

生态学分类的环境解释涉及到远多于象 SDA 和 MDA 这样的多变量过程。生态学家必需深思这样一个普遍存在的(通常是模棱两可的)问题,即在多变量分析中应该包括哪些环境因子。这些分析中是否只能包括那些有可能作用显著的因子,或许该判断应该基于一些以前的单变量检验?或者只包括那些起作用的变量(Spomer 1973),即那些和生物有直接相互作用(能量或质量)的因子(如水或养分)?还是选择包括更多的易于测定而不起作用的变量(如,地点的海拔、坡向、坡度)——即间接影响起作用变量的变量?

由于包括了大量的变量(其中很可能包括一些“弱”因子),实际上,判别分析可能被减弱(见 23.4 部分中的示例)。而且,使用大量的变量可能要冒包含有高度相关因子之险,这经常给多变量统计过程带来问题(见 Sokal 和 Rohlf 1981)。例如, 23.7 部分中假设的不相关变量

(Sneath 和 Sokal 1973)。生态学因子之间的相互作用是预料之中的，而不是一个例外，在较简单的单变量检验之后，利用多变量统计检验的原因之一是得到一个相互作用存在于何处的更好解释（正如 SDA 中使用的协方差）。

生态学家必须在选择（和测定）多变量分析中的所有生态因子时十分小心。由于这个原因，我们不推荐使用逐步 SDA 和逐步 MDA，因为它们具有一种把所有测定的变量（重要或不重要）都“抛入”到 MDA 中，让逐步过程选择那些“统计学上”重要的变量的倾向；而这些变量中可能有许多没有什么生态学意义的变量。在选择环境因子时，生态学家如果能采用很好的判断方法，那么，SDA 和 MDA 可以成为解释群落变异格局的有力工具。

23.7 总结和推荐

1. 当从一个分类中发现了 SU 聚类的有趣格局时，下一步该做的是对此格局提出一个环境解释。如果群落聚类对（SU 组对）之间环境因子（加在一起）存在显著差异，那么简单判别分析（SDA）可以成为一种有力的多变量统计方法。SDA 还提供了有关每一个环境因子在判别过程中相对重要性的信息（23.1 部分）。

2. 对于那些对同时比较三个或更多群体组（而不是象 SDA 中一对一的）感兴趣的生态学家，我们推荐他们使用复合判别分析（MDA，也称为经典变量分析）。

3. 为了说明判别分析的过程和计算这一简单目的，我们采用了小的数据集。然而，当同学们把这些方法应用于他们自己的数据时，我们提醒你们要小心注意基本参数和多变量假设。样本大小，尤其是每组中 SU 的数量是很重要的。如果样本小，基本的假设就不可能维持，那么我们推荐使用非参数的多变量方法。

4. 对于 SDA，在分析中包含“弱”判别变量可以减小分析的能力（见 23.4 部分示例）。因此，我们建议，在选择 SDA 和 MDA 中的环境因子时要非常小心。具有高度生态学意义的，而不是纯统计学意义的因子，才应该被采用。我们不提倡使用逐步 SDA 或逐步 MDA（23.6 部分）。

第二十四章 排序的解释

在这一章里，我们将提供建立 SU 排序坐标(由种数据导出)和环境因子关系的各种方案，这些排序坐标可以概念化为与潜在环境梯度有关的合成轴。我们将阐述用以检验这些关系的一元和多元线性回归的应用。

24.1 基本方法

植物生态学的一个中心议题是阐明植被与环境因子的关系，这导致了一组技术即直接和间接的梯度分析产生(定义见 17 章)。解释间接排序的方案是非常不同的。Bray 和 Curtis 在一个极点排序中(PO)利用了 SU_s 内种重要性和环境因子测量值的简单图。一些人对于种和环

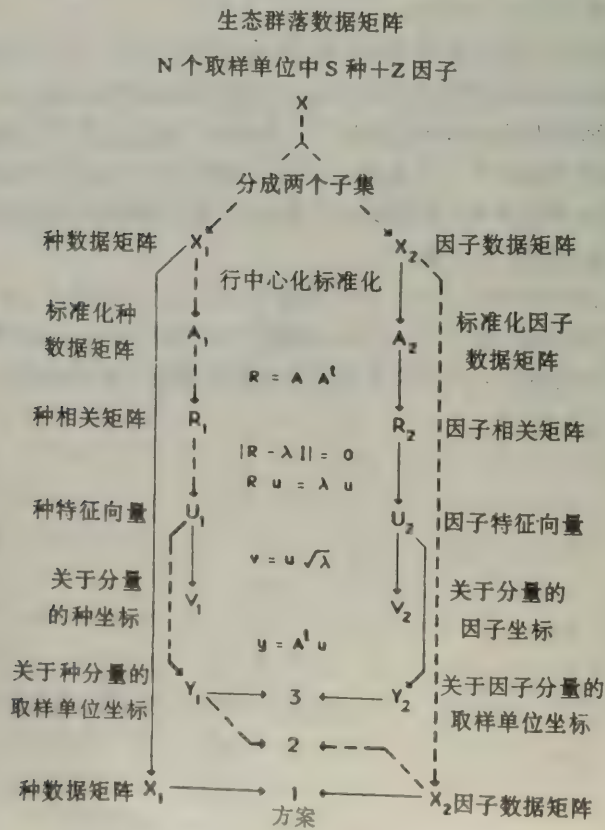


图 24.1 使种—多度或合成种分量(来自 PCA、COA 或 DPC)与环境因子(直接梯度)或环境分量(间接合成梯度)相关的方案，本章中阐述的方案是 2(虚线)，矩阵记法遵照 19 章的

境因子混合的数据矩阵试用了主分量分析(PCA)(Barbham 和 Norris 1970, Walker 和 Wehrhahn 1971)。数学技术的一些变体也得到了应用,包括典范相关分析(Canonical correlation analysis)(Gauch 和 Wentworth 1976)、对应分析(Correspondence analysis, COA)和剩余排序排序(Residual ordination analysis)(Carleton 1984)。

图 24.1 阐述了一个与排序环境解释有关而发展的基本模型。回忆一个群落数据矩阵有两个变量子集,一个是种,另一个是环境因子(如图 22.1)。对于给定的这两个子集,给出了三种方案,所有这些都可能包含相关和回归分析的使用,最直接的方法(方案 1)是利用相关或回归分析,通过 N 个 SUs 使每个种与每个环境因子联系起来。方案 2 是,利用种数据特征分析导出 SUs 坐标(如 PCA)并将这些坐标回归到在这些 SUs 中测定的环境因子上。从种和因子数据的特征分析中导出 SU 排序坐标并建立这些结果间的关系(方案 3)也是可能的

方案 1 是直接排序的例子,而方案 2、3 是间接排序的例子(Austin 等 1984, Gauch 1982, Whittaker 1967),在这一章里我们将阐述方案 2(如图 24.1 中虚线所示)。

24.2 步骤

下面步骤阐述方案 2

步骤 1. 获得 SU 坐标。获得 SU 在 PCA 或 COA (或 DPC)排序轴上的坐标,注意由图 24.1, SU 坐标是由种相关矩阵 R 获得的(如果来自 DPC 的结果表明某组数据有明显的非线性结构,那么就要利用这个“展开”DPC 轴,而不是(或一起)用 PCA 的前两轴。

步骤 2. 回归分析。计算使 SU 坐标与在每个 SU 中环境变量建立起关系的一元和多元线性回归系数。

a. 简单的线性回归模型(SLR)是:

$$Y = B_0 + B_1X \tag{24.1}$$

Y 代表 SU 排序坐标, X 在 SUs 中测定的感兴趣的特殊环境因子值。公式(24.1)包括使用最小二乘法技术(Least-squares techniques)(见 Sokal 和 Rohlf 1981)。首先我们计算平均数 \bar{y} 和 \bar{x} , Y 和 X 的校正平方和 $\sum y^2$ 和 $\sum x^2$, X 和 Y 间的校正交叉乘积和 $\sum xy$,而后由下式计算系数 B_1 (斜率)

$$B_1 = \frac{\sum xy}{\sum x^2} \tag{24.2}$$

和 B_0 (截距):

$$B_0 = \bar{y} - (B_1\bar{x}) \tag{24.3}$$

另一个感兴趣的是测定系数 r^2 , 它是回归 Y 到 X 上变差与总变差之比的估计值。

$$r^2 = \frac{(\sum xy)^2 / \sum x^2}{\sum y^2} \tag{24.4}$$

SLR 的统计显著性由 r^2 的平方根(即 r)测定并将它与 $P=0.05$, 自由度 $N-2$ 的概率表相比较(如 Steele 和 Torrie 1960, 表 A.13),或测定 B_1 的 T 统计量(见 Sokal 和 Rohlf 1981, 16.2 节)。

b. 多元线性回归模型是

$$Y = b_0 + (b_1X_1) + (b_2X_2) + \cdots + (b_pX_p) \quad (24.5)$$

Y 定义如上[公式(24.1)]。X 代表在 SUs 中测定的 P 个环境因子值, 计算 Y 的平均数(\bar{y}), 每个 X 的平均数(即 $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p$), Y 的总的平方校正和($\sum y^2$), 每个 X 的总的平方校正和(如 $\sum x_1^2$, X 与 Y 的($\sum x_1y$)和 X 之间的(如 $\sum x_1x_2$)校正交叉乘积和, 通过解线性正规方程组确定偏回归系数 b's(Sokal 和 Rohlf 1981), 利用计算机化的算法这个正规方程组很容易解, 如果我们举 2 个未知数(两个环境因子)2 个方程的简单情形, 这两个方程是:

$$b_1 \sum x_1^2 + b_2 \sum x_1x_2 = \sum x_1y \quad (24.6a)$$

$$b_1 \sum x_2x_1 + b_2 \sum x_2^2 = \sum x_2y \quad (24.6b)$$

系数 b_1, b_2 可通过下式解出

$$b_1 = \frac{(\sum x_2^2 \sum x_1y) - (\sum x_1x_2 \sum x_2y)}{V} \quad (24.7a)$$

$$b_2 = \frac{(\sum x_1^2 \sum x_2y) - (\sum x_2x_1 \sum x_1y)}{V} \quad (24.7b)$$

$$V = (\sum x_1^2 \sum x_2^2) - (\sum x_1x_2 \sum x_2x_1), \text{ 和} \quad (24.7c)$$

$$b_0 = \bar{y} - [(b_1\bar{x}_1) + (b_2\bar{x}_2)]$$

多元测定系数 R^2 是

$$R^2 = \frac{(b_1 \sum x_1y) + (b_2 \sum x_2y)}{\sum y^2} \quad (24.8)$$

它是在 Y 中 X 共同占有的离差成分估计值, MLR 的统计显著性由 F 比率(Sokal 和 Rohlf 1981, 16.2 节)测定。

步骤 3. 检验结果。通过检验回归系数的方向(符号)和统计显著性, 解释所得到的结果, 这将指出哪个排序分量同哪个环境因子(梯度)建立起了关系。

必须强调的一点是, 即使很强的统计相关或回归也许能或不能代表有意义的生态相关; 生态学家应该十分仔细地对待这些结果, 对统计结果所表示的可能关系进行追加的讨论, 同时也要记住排序轴是合成的(它们由 PCA 或 DPC 导出的), 也许能或不能与有意义的潜在生态趋势相对应。

24.3 举例: 计算

我们利用 2 个种 2 个环境因子在 5 个 SUs(表 23.2)中的模拟数据阐述了排序的环境解释的计算, SU 在 PCA 的主分量 I 和 II(图 19.4)和无偏主分量(图 21.2)的得分将被利用, 同时被利用的还有环境因子 X_1 和 X_2 , 这个简单的数据对于阐述方案 2(图 24.1)的计算是足够的。

步骤 1. 获得 SU 坐标。这些结果总结在表 24.1 中。

步骤 2. 回归分析。我们现在可以获得排序坐标回归到环境因子上的一元或多元回归系数。为阐述这个计算，我们选择了 DPC(I) 坐标(表 24. 1)作为在公式(24. 1)和(24. 5)中的因变量 Y。

a. 一元线性回归:给定 $Y=Y_3$ [即 DPC(I)]和 $X=X_1$ (表 24. 1)，斜率和截距系数按公式(24. 2)和(24. 3)计算:

$$B_1 = -3.064/5.2 = -0.589$$
$$B_0 = 0.026 - (-0.589) \times 2.8 = 1.675$$

表 24. 1 5 个 SU_s 在主分量(PC) I 和 II 和在无偏主分量(DPC)上的作标, 以及 2 个环境因子 X₁、X₂ 在这 5 个 SU_s 中的值。同时也给出了这些分量和因子的平均数, 校正平方和, 校正交叉乘积和

SU _s	分量			环境因子		
	PC(I)	PC(II)	DPC(I)	X ₁	X ₂	
	Y ₁	Y ₂	Y ₃			
(1)	-0.62	-0.28	-0.58	4.2	6.1	
(2)	+0.56	-0.12	+0.71	1.6	7.7	
(3)	+0.78	+0.12	+1.00	2.4	8.3	
(4)	0.00	-0.50	0.00	2.0	8.0	
(5)	-0.72	+0.28	-1.00	3.8	5.9	
$\bar{y} =$	0.00	-0.10	0.026	$\bar{x} =$	2.8	7.2
$\sum y^2 =$	1.825	0.386	2.837	$\sum x^2 =$	5.2	5.0
$\sum x_1 y =$	-2.572	0.384	-3.064	$\sum x_1 x_2 =$	-4.520	
$\sum x_2 y =$	2.756	-0.384	3.393	$\sum x_2 x_1 =$	$\sum x_1 x_2$	

测定系数[公式(24. 4)]

$$r^2 = [(-3.064) \times (-3.064)/5.2]/2.873 = 0.636$$

由它给出 $r=0.798$, 这个值在 $P=0.05(df=3)$ 时不显著。

对于环境因子 X₂

$$B_1 = 3.393/5.0 = 0.679$$
$$B_0 = 0.026 - 0.679 \times 7.2 = -4.862$$

测定系数:

$$r^2 = [3.393 \times 3.393/5.0]/2.837 = 0.812$$

由它给出 $r=0.90$, 这个值在 $P=0.05(df=3)$ 时是显著的。

最后, DPC 得分与环境因子 X₂ 是显著相关的。SU_s 沿着以这些种多度的所有离差的无偏

主分量位置 and 这些离差与潜在的 X_2 梯度呈显著相关。Y (DPC (I) 得分) 对应于 X_2 的图阐述了这种关系 (图 24.2)。

b. 多元线性回归 Y=DPC (I) 得分的斜率和截距系数, 首先由解方程 (24.6) 确定。

$$b_1 \times 5.2 + b_2 \times (-4.520) = -3.064$$

$$b_1 \times (-4.520) + b_2 \times 5.0 = 3.393$$

由公式 (24.7)

$$b_1 = [5.0 \times (-3.064) - (-4.520) \times 3.393] / 5.570 = +0.0029$$

$$b_2 = [5.2 \times 3.393 - (-4.520) \times (-3.064)] / 5.570 = 0.681$$

$$V = 5.2 \times 5.0 - (-4.520) \times (-4.520) = 5.570$$

$$b_0 = 0.026 - [0.0029 \times 2.8 + 0.681 \times 7.2] = -4.869$$

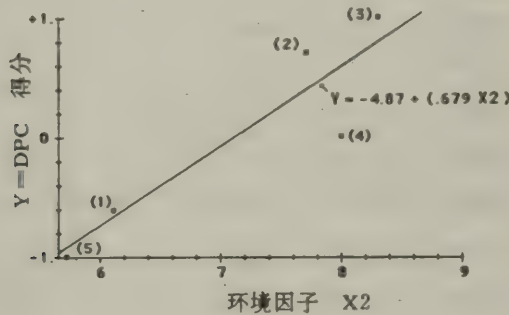


图 24.2 SU_s 在无偏主分量上的得分与环境因子 X_2 线性回归关系

由公式 (24.8) 多元测定系数为

$$R^2 = [(-0.0029) \times (-3.064) + 0.681 \times 3.393] / 2.837 = 0.818$$

它给出 $R = 0.904$ 在 $P = 0.05$ 时不显著, 因为自由度较低 ($df = 2$)。

纵然在这一简单的数据组中 MLR 统计上并不显著, 我们还是要阐述一下每个因子对总的限定系数 (R^2) 贡献的相对比百分数是怎样被测定的。使用与在 SDA 中因子对于 Mahalanobis 距离 (D^2) 的相对贡献相似类的计算 (23 章)。因子 X_1 和 X_2 的相对贡献是:

$$RC_1 = [(-0.0029) \times (-3.064) / 2.837 / 0.818] \times 100\% = 0.4\%$$

$$RC_2 = (0.681 \times 3.393 / 2.837 / 0.818) \times 100\% = 99.6\%$$

由此可见, 对于解释 SU 沿着 DP (I) 放置产生的离差 X_2 因子比 X_1 更重要。

24.4 例子: 巴拿马蟑螂

在表 23.4 巴拿马六个发生地 (SU_s) 的数据包括了两个因子, 高度和降雨的一个子集。利用这些数据和 PCA 的结果 (表 19.4), 运行 BASIC 程序 PCREG. BAS 进行 SLR 和 MLR 分析。[因为 DPC 对这组数据分析的结果不显著 (表 21.2), 所以我们没有包括无偏主分量]。由

PCREG. BAS 输出的 PC I 和 PC I 回归到因子 1 和因子 2 上的结果表明,分量和因子之间没有显著的线性关系。 r^2 (SRL)和 (R^2) 都很低且不显著。

表 24.2 巴拿马 6 个发生地分量得分到两个环境因子高度(X_1)和降水(X_2)上的一元和多元线性回归

环境因子和主分量统计量							
变量		平均数	方差	标准差	标准误		
因子 X1		0.461	0.339	0.582	0.238		
因子 X2		2.833	1.367	1.169	0.477		
分量 I		-0.000	0.605	0.778	0.318		
分量 II		-0.000	0.235	0.485	0.198		
一元线性回归							
PC	比较 因子	回归系数		r ²	B ₁ 的 标准误	t 统计量	df
		B ₀	B ₁				
I	X ₁	-0.261	0.566	0.180	0.605	0.936	4
I	X ₂	-0.611	0.216	0.105	0.315	0.685	4
II	X ₁	0.071	-0.154	0.034	0.409	-0.377	4
II	X ₂	-0.131	0.046	0.012	0.206	0.225	4
多元线性回归							
分量 I : $Y = -0.523 + (0.466X_1) + (0.109X_2)$							
分量 II : $Y = -0.178 + (-0.250X_1) + (0.104X_2)$							
多元测定系数、F 统计量和 df							
		R ²	F	df			
分量 I		0.201	0.388	2,3			
分量 II		0.083	0.136	2,3			

24.5 例子:威斯康星森林

在表 23.6 给出了两个因子,土壤质地、“立地动态”指数,它们与 10 个高地威斯康星森林种的数据相对应。利用 PCA(表 19.5)和 DPC(表 21.4)的结果,在表 24.3 中给出 PCREG. BAS 程序 SLR 和 MLR 分析的结果。注意 PC(I)和 DPC 与因子 X_2 (立地动态指数)呈负相关,与因子 X_1 (土壤质地)呈正相关。

测定系数高,t 和 F 统计量十分显著($P < 0.005$),如预料的那样,经 SLR 和 MLR 分析,PC(I)与这两个因子的关系不显著,因为它代表一个由 DCP 分析有效移走的一个弓形。

在 PC(I)和 DPC 上 SU 的作标对应于因子 1(图 24.3a)和因子 2(图 24.3b)的图明显地表

示了这些线性相关。

表 24.3 来自主分量 PC(I)和 PC(II)和无偏主分量(DPC III)到两个因子, 土壤质地(X₁)和立地动态指数上的一元、多元线性回归分析的结果

环境因子和主分量统计量						
变量	平均数		方差	标准差	标准误	
因子 X1	2.100		2.100	1.449	0.458	
因子 X2	2.700		2.456	1.567	0.496	
分量 PC(I)	0.000		0.492	0.702	0.222	
分量 PC(II)	0.000		0.131	0.363	0.115	
分量 DPC	0.074		0.620	0.788	0.249	
一元线性回归						
比较	回归系数			B ₁ 的		
变量	B ₀	B ₁	r ²	标准误	t 统计量	自由度 df
PC(I),X ₁	-0.729	0.347	0.514	0.119	2.907	8
PC(I),X ₂	-0.729	0.347	0.514	0.119	2.907	8
PC(II),X ₂	1.082	-0.401	0.801	0.070	-5.675	8
PC(II),X ₁	-0.283	0.135	0.291	0.074	1.813	8
PC(II),X ₂	-0.058	0.021	0.08	0.081	0.266	8
DPC,X ₁	-0.818	0.425	0.610	0.120	3.540	8
DPC,X ₂	1.275	-0.445	0.783	0.082	-5.369	8
多元线性分析						
变量	多元回归系数			R ²	F	df
	b ₀	b ₁	b ₂			
PC(I)	0.563	0.147	-0.323	0.863	22.07	2,7
PC(II)	-0.838	0.221	0.139	0.532	3.98	2,7
DPC	0.491	0.222	-0.327	0.895	29.86	2,7

24.6 排序解释的附加讨论

这一章里我们提供了方案 2(图 24.1), 它把由一个排序(如 PCA 和 DPC)导出的分量(坐标)直接回归到非生物(环境)因子的任何一个(SLR)或一起(MLR)上。种多度数据被用来产生那些 SUs 分量, 之后, 它们被与环境因子有关的数据子集检验。我们现在将描述一下其它的方案, 在这些方案中种和环境数据同时被分析以阐述它们之间的关系。

首先把生物和非生物的群落数据子集处理为一个大的数据集, 并使这个矩阵隶属于例如一个 PCA, 以检验种和环境因子 SU 坐标的对应性(相对位置), 这看上去是比较直观的, 但由 PCA 提炼出关于这种类型混合数据的分量不再是生物成分, 而生物-非生物混合成分。生态上, 其目的就是提取生物成分并对它们作关于非生物影响的解释, 这就变得困难了, 由于

它是混合了的生物—非生物分量，因而这些结果受种间相关、因子间相关、种和因子间相关的强烈影响。事实上由于种和环境因子不是以同一单位测定的，故这些数据必须以某种方式标准化。

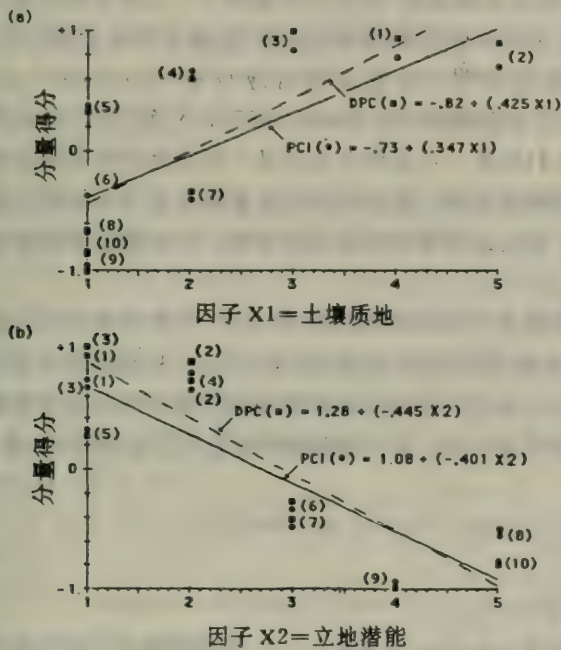


图 24.3 10 个高地威斯康星森林立地在头一个主分量和无偏分量(DPC)上得分对应于 (a)因子 X₁—土壤质地, 和(b)因子 X₂—立地动态指数图

Walker 和 Wehrhahn(1971)描述了用 PCA 对于种—因子结合的数据分析一个过程，以使提出的分量反映在生物子集中的离差和非生物子集对分量自然性不是过度地影响。他们的过程基本上包括：通过除以每个因子的最大值调节这些因子在 0—1 的范围内使非生物数据标准化，然后在代入 PCA 之前再进一步把这些数据变为非常小的值(通过除以 1000)。结果就是这些调节了的非生物因子总的方差—协方差的贡献相对于生物变量是小的。因此，提取出的分量将在很大程度的上反映生物变量的成分，之后非生物因子的坐标可被调节以提供它们与生物分量之间的关系。Walker 和 Wehrhahn 使用了这个过程对于加拿大湿地数据进行了分析获得了可以解释的结论。

典范相关分析(Canonical correlation analysis) (Gauch 和 Wentworth)是同时分析混合的种和环境因子数据的另一个过程，这个多元的统计过程使用带有多元自变量和多元因变量的线性典范方程(与多元线性回归相比，多元线性回归仅用多元自变量)。确定偏典范分析系数以使因变量(种)子集与自变量子集(环境因子)的相关最大化，在这个过程中，由于种和因子坐标可被画在同一个典范排序系统中，而且用它们之间的位置定义它们之间的相互关系，这在直观上看，对于生态解释是理想的，但典范分析对于生态数据的许多应用产生了有限的可利用结果(如 Austin 1968)。它对于种—多度关系和因子梯度的线性假定是相当严格的(Gauch

和 Wentworth 1970)。当然本章提供的线性回归方程也有类似的假设,下面将讨论非线性回归。

典范对应分析(Canonical correspondence analysis)是一个新的过程,它把对于种—立地对应分析的相互平均方法(结合无偏选择)和对于环境因子—立地的加权多元回归分析结合为一个算法(Ter Braak 1986, 1987),这个典范对应分析(无偏化和非无偏化的)与无偏对应分析对于三个数据集的处理比较表明,无偏典范对应分析(Detrended canonical correspondence analysis)非常有效地提供了生态解释(Ter Braak 1986),尽管这种方法的详细算法不在本书范围之内,但我们建议学生们注意一下这种方法的进一步应用和对这种方法的检验。Whittaker (1987)描述了一种用无偏对应分析,秩相关和非度量多维调节成功地识别植被—环境因子复合体的四个步骤,另一个新的过程就是剩余排序分析,它主要是强调解释关于排序分量的剩余离差(Carleton 1984)。

尽管本章的处理仅限定在一元和多元线性回归上,但各种曲线回归也可被利用以确定单一的分量(如 PC I)与单的环境因子(如湿度)之间的关系。从相对简单的指数函数、幂函数(如 Legendre 和 Legendre 1983)到相对复杂的高斯模型这些曲线回归模型都是可以利用的。为了将一个分量与多个因子联系在一起,多元曲线模型也是可以被利用的(如多项式回归 Sokal 和 Rohlf 1981)。

24.7 总结和推荐

1. 当寻找沿一排序轴感兴趣的 SUs 格局的一个环境解释时,用线性回归使在这个轴上的 SU 坐标与环境变量建立关系的方案可以提供有意义的生态解释(24.1 节),此时所供的梯度应是狭窄的,种的反应是线性的(图 21.1a)。

2. 一元和多元线性回归均可被用来探索排序轴和环境因子之间的关系,在两种方法都推荐的同时,但多元回归可能更有信息,因为在自然界中,群落的数据是多元的,我们提醒学生,统计显著性的回归也许能或不能含有有意义的生态关系(24.2 节)。

3. 我们不推荐用主分量分析处理种和环境变量结合在一起的一个大的数据组。由这种 PCA 获得的分量将是生物—非生物混合分量,对于这些立地因子作种离差的生态解释是不清楚的。尽管立地的因子可被调节为非常小的值,因而对主分量的贡献也小,为了同时分析种和环境的关系,这种特殊设计的方法的使用才被选择(24.6 节)。

4. 典范对应分析是一个新的步骤,它对于同时探索立地种—因子数据关系看来是比较有前途的,它是把对种数据的无偏对应分析和对环境数据加权的多元回归结合在一起的一种算法,这种技术的进一步发展可能产生一个有力的生态工具。

参考文献

- Abrams, P. A. (1980). Some comments on measuring niche overlap. *Ecology* 61:44—49
- Alatalo, R. V. (1981). Problems in the measurement of evenness in ecology. *Oikos* 37:199—204.
- Alatalo, R. V. and R. Alatalo (1977). Components of diversity; multivariate analysis with interaction. *Ecology* 58:900—906.
- Able, K. P. and B. R. Noon (1976). Avian community structure along elevational gradients in the northeastern United States. *Oecologia* 26:275—294.
- Anderberg, M. R. (1973). *Cluster Analysis for Applications*. Academic Press, New York.
- Anderson, A. J. B. (1971). Ordination methods in ecology. *Journal of Ecology* 59:713—726.
- Anderson, D. J. (1965). Classification and ordination in vegetation science; controversy over a nonexistent problem? *Journal of Ecology* 53:521—526.
- Andrews, M. J. and D. Richard (1980). Rehabilitation of the Inner Thames estuary. *Marine Pollution Bulletin* 11:327—331.
- Archie, J. W. (1984). A new look at the predictive value of numerical classification. *Systematic Zoology* 33:30—51.
- Ashton, D. H. (1976). The vegetation of Mount Piper, Central Victoria; A study of a continuum. *Journal of Ecology* 64:463—483.
- Austin, M. P. (1968). An ordination study of a chalk grassland community. *Journal of Ecology* 56:739—757.
- Austin, M. P. (1985). Continuum concept, ordination methods and niche theory. *Annual Review of Ecology and Systematics* 16:39—61.
- Austin, M. P. and L. Orloci (1966). Geometric models in ecology. II. An evaluation of some ordination methods. *Journal of Ecology* 55:217—227.
- Austin, M. P., R. B. Cunningham, and P. M. Fleming (1984). New approaches to direct gradient analysis using environmental scaling and statistical curve—fitting procedures. *Vegetatio* 55:11—28.
- Bannister, P. (1968). An evaluation of some procedures used in simple ordinations. *Journal of Ecology* 54:665—674.
- Barkham, J. P. and J. M. Norris (1970). Multivariate procedures in an investigation of vegetation and soil relations of two beech woodlands, Cotswold Hills, England. *Ecology* 51:630—639.
- Bartlett, M. S. (1954). A note on the multiplying factors for various chi—squared approximations. *Journal Royal Statistical Society, Series B* 16:296—298.
- Beals, E. W. (1960). Forest bird communities in the Apostle Islands of Wisconsin. *Wilson Bulletin* 72:156—181.
- Beals, E. W. (1965). Ordination of some corticolous cryptogamic communities of south—central

Wisconsin. *Oikos* 16:1—8.

- Beals, E. W. (1984). Bray—Curtis ordination: An effective strategy for analysis of multivariate ecological data. *Advances in Ecological Research* 14:1—55.
- Bennet, S. H., J. Whitfield Gibbons, and J. Glanville (1980). Terrestrial activity, abundance and diversity of amphibians in differently managed forest types. *American Midland Naturalist* 103:412—416.
- Bliss, C. I. and D. W. Calhoun (1954). *An Outline of Biometry*. Yale Co—op Publ., New Haven, CT.
- Bliss, C. I. and R. A. Fisher (1953). Fitting the binomial distribution to biological data and a note on the efficient fitting of the negative binomial. *Biometrics* 9:176—200.
- Bloom, S. A. (1981). Similarity indices in community studies: Potential pitfalls. *Marine Ecology* 5:125—128.
- Boesch, D. F. (1977). Application of Numerical Classification in Ecological Investigations of Water Pollution. EPA—600/3—77—033. US Environmental Protection Agency, Corvallis, OR.
- Bradfield, G. E. and N. C. Kenkel (1987). Nonlinear ordination using flexible shortest path adjustment of ecological distances. *Ecology* 68:750—753.
- Bray, J. R. and J. T. Curtis (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs* 27:325—349.
- Brown, M. J., D. A. Ratkowsky and P. R. Minchin (1984). A comparison of detrended correspondence analysis and principal co—ordinates analysis using four sets of Tasmanian vegetation data. *Australian Journal of Ecology* 9:273—279.
- Brown, R. T. and J. T. Curtis (1952). The upland conifer—hardwood forests of northern Wisconsin. *Ecological Monographs* 22:217—234.
- Bultman, T. L. and S. H. Faeth (1985). Patterns of intra— and interspecific associations in leaf—mining insects on three oak host species. *Ecological Entomology* 10:121—129.
- Campbell, B. M. (1978). Similarity coefficients for classifying relevés. *Vegetatio* 37:101—109.
- Carleton, T. J. (1980). Non—centered component analysis of vegetation data: a comparison of orthogonal and oblique rotation. *Vegetatio* 42:59—66.
- Carleton, T. J. (1984). Residual ordination analysis: a method for exploring vegetation—environment relationships. *Ecology* 65:469—477.
- Carpenter, S. R. and J. E. Chaney (1983). Scale of spatial pattern: four methods compared. *Vegetatio* 53:153—160.
- Cattell, R. B. (1952). *Factor Analysis*. Harper, New York.
- Chang, D. H. S. and H. G. Cauch, Jr. (1986). Multivariate analysis of plant communities and environmental factors in Ngari, Tibet. *Ecology* 67:1568—1575.
- Chardy, P., M. Glemarec, and A. Laurec (1976). Application of inertia methods to benthic marine ecology: practical implications of the basic options. *Estuary and Coastal Marine Science*

4;179—205.

- Christensen, N. L. (1977). Changes in structure, pattern and diversity associated with climax forest maturation in Piedmont, North Carolina. *American Midland Naturalist* 97;176—188.
- Clapham, A. R. (1936). Over—dispersion in grassland communities and the use of statistical methods in plant ecology. *Journal of Ecology* 24;232—251.
- Claridge, M. F. and M. R. Wilson (1981). Host plant association, diversity and species—area relationships of mesophyll—feeding leafhoppers of trees and shrubs in Britain. *Environmental Entomology* 6;217—238.
- Clark, P. J. and F. C. Evans (1954). Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* 35;445—453.
- Clifford, H. T. and W. Stephenson (1975). *An Introduction to Numerical Classification*. Academic Press, New York, pp. 37—46.
- Clifford, H. T. and W. T. Williams (1976). Similarity measures. In *pattern Analysis in Agricultural Science* (W. T. Williams, Ed.). Elsevier, New York.
- Clymo, R. S. (1980). Preliminary survey of the peat—bog Hummell Knowe Moss using various numerical methods. *Vegetatio* 42;129—148.
- Cochran, W. G. (1963). *Sampling Techniques*, 2nd ed. Wiley, New York.
- Cody, M. L. (1974). *Competition and the Structure of Bird Communities*. Princeton University Press, Princeton, NJ.
- Coetsee, B. J. and M. J. A. Werger (1975). On association analysis and the classification of plant communities. *Vegetatio* 30;201—206.
- Connell, J. H. (1963). Territorial behaviour and dispersion in some marine invertebrates. *Research in Population Ecology* 5;87—101.
- Conner, R. N. and C. S. Adkinson (1976). Discriminate function analysis; A possible aid in determining the impact of forest management on woodpecker nesting habitat. *Forest Science* 22;122—127.
- Cook, C. W. and R. Hurst (1963). A quantitative measure of plant association in ranges in good and poor condition. *Journal of Range Management* 15;266—274.
- Cooley, W. W. and P. R. Lohnes (1971). *Multivariate Data Analysis*. Wiley, New York.
- Cottam, G. and J. T. Curtis (1956). The use of distance measures in phytosociological sampling. *Ecology* 37;451—460.
- Cottam, G. , F. G. Goff, and R. H. Whittaker (1978). Wisconsin comparative ordination. In *Ordination of Plant Communities* (R. H. Whittaker, Ed.). W. Junk , The Hague , pp. 185—213.
- Cox, G. W. (1985). *Laboratory Manual of General Ecology*, 5th ed. Brown, Dubuque, IA.
- Cox, T. F. and T. Lewis (1976). A conditioned distance ratio method for analyzing spatial patterns. *Biometrika* 63;483—491.
- Crawford, R. M. M. and D. Wishart (1967). A rapid multivariate method for the detection

- and classification of groups of ecologically related species. *Journal of Ecology* 55:505—524.
- Crawford, R. M. M. and D. Wishart (1968). A rapid classification and ordination method and its application to vegetation mapping. *Journal of Ecology* 56:385—404.
- Crow, T. R. and D. F. Grigal (1979). A numerical analysis of arborescent communities in the rain forest of the Luquillo Mountains, Puerto Rico. *Vegetatio* 40:135—146.
- Culp, J. M. and R. W. Davies (1980). Reciprocal averaging and polar ordination as techniques for analyzing lotic macroinvertebrate communities. *Canadian Fishery and Aquatic Science* 37:1358—1364.
- Curtis, J. T. and R. P. McIntosh (1951). The upland forest continuum in the prairie—forest border region of Wisconsin. *Ecology* 32:476—496.
- Dale, M. B. and H. T. Clifford (1976). On the effectiveness of higher taxonomic ranks for vegetation analysis. *Australian Journal of Ecology* 1:37—62.
- Danin, A. (1976). Plant species diversity under desert conditions I. Annual species diversity in the Dead Sea Valley. *Oecologia* 22:251—259.
- Dargie, T. C. D. (1984). On the integrated interpretation of indirect site ordinations: a case study using semi-arid vegetation in southeastern Spain. *Vegetatio* 55:37—55.
- David, F. N. and P. G. Moore (1954). Notes on contagious distributions in plant populations. *Annals of Botany* 18:47—53.
- De Jong, T. M. (1975). A comparison of three diversity indices based on their components of richness and evenness. *Oikos* 26:222—227.
- De pablo, C. L., B. Peco, E. F. Galiano, J. P. Nicolas, and F. D. Pineda (1982). Space—time variability in Mediterranean pastures analyzed with diversity parameters. *Vegetatio* 50:113—125.
- Del Moral, R. (1972). Diversity patterns in forest vegetation of the Wenatchee Mountains, Washington. *Bulletin Torrey Botanical Club* 99:57—64.
- Dice, L. R. (1945). Measures of the amount of ecological association between species. *Ecology* 26:297—302.
- Dice, L. R. (1952). Measure of the spacing between individuals within a population. *Contribution Laboratory Vertebrate Biology, University of Michigan, East Lansing, MI.*
- Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, New York.
- Diggle, P. J., J. Besag, and J. T. Gleaves (1976). Statistical analysis of spatial point patterns by means of distance methods. *Biometrics* 32:659—667.
- Dix, R. L. and J. E. Butler (1960). A phytosociological study of a small prairie in Wisconsin. *Ecology* 41:316—327.
- Dixon, W. J. and M. B. Brown (Eds) (1979). *BMDP—79: Biomedical Computer Programs, P—Series*. University of California Press, Berkeley, CA.
- Doncaster, C. P. (1981). The spatial distribution of ant nests on Ramsey Island, South Wales. *Journal of Animal Ecology* 50:195—218.

- Drake, J. A. (1984). Species aggregation: the influence of detritus in a benthic invertebrate community. *Hydrobiology* 112:109—115.
- Duncan, T. and G. F. Estabrook (1976). An operational method for evaluating classifications. *Systematic Botany* 1:373—382.
- Eberhardt, L. L. (1967). Some developments in distance sampling. *Biometrics* 23:207—216.
- Eberhardt, L. L. (1976). Quantitative ecology and impact assessment. *Journal of Environmental Management* 4:27—70.
- Elliott, J. M. (1973). Some methods for the Statistical Analysis of Samples of Benthic Invertebrates. Scientific Publication No. 25, Freshwater Biological Association, Ambleside, Westmorland, Great Britain.
- Erman, D. C. (1973). Ordination of some littoral benthic communities in Bear Lake Utah—Idaho. *Oecologia* 13:211—226.
- Everitt, B. (1974). *Cluster Analysis*. Wiley, New York.
- Faith, D. P. (1983). Asymmetric binary similarity measures. *Oecologia* 57:287—290.
- Faith, D. P. (1984). Patterns of sensitivity of association measures in numerical taxonomy. *Mathematical Biosciences* 69:199—207.
- Faith, D. P., M. P. Austin, L. Belbin, and C. R. Margules (1985). Numerical classifications of profile attributes in environmental studies. *Journal of Environmental Management* 20:73—85.
- Faith, D. P., P. R. Minchin, and L. Belbin (1987). Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 69:57—68.
- Fasham, M. J. R. (1977). A comparison of nonmetric multidimensional scaling, principal components and reciprocal averaging for the ordination of simulated coenoclines, and coenoplanes. *Ecology* 58:551—561.
- Feoli, E. and M. Iacono (1983). A resemblance function based on probability: Applications to field and simulated data. *Vegetatio* 53:3—9.
- Fewster, P. H. and L. Orloci (1983). On choosing a resemblance measure for non—linear predictive ordination. *Vegetatio* 54:27—35.
- Findley, J. S. and M. T. Findley (1985). A search for pattern in butterfly fish communities. *American Naturalist* 126:800—816.
- Fish, E. B. (1976). Comparisons of phytosociological methods of classification on a desert grassland site. *Agro—Ecosystems* 2:173—194.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Human Genetics* 7:179—188.
- Foran, B. D., G. Bastin, and K. A. Shaw (1986). Range assessment and monitoring arid lands: the use of classification and ordination in range survey. *Journal Environmental Management* 22:67—84.
- Forsythe, W. L. and O. L. Loucks (1972). A transformation for species response to habitat

- factors. *Ecology* 53:1112—1119.
- Garratt, M. W. and R. K. Steinhorst (1976). Testing for significance of Morisita's, Horn's and related measures of overlap. *American Midland Naturalist* 96:245—251.
- Gauch, H. G. (1973). The relationship between sample similarity and ecological distance. *Ecology* 54:618—622.
- Gauch, H. G. (1977). ORDIFLEX—A Flexible Computer Program for Four Ordination Techniques: Weighted Averaging, Polar ordination, Principal Component Analysis and Reciprocal Averaging. Release B, Cornell University. Ithaca, NY.
- Gauch, H. G. (1982). *Multivariate Analysis in Community Ecology*. Cambridge University Press, New York.
- Gauch, H. G. and W. M. Scruggs (1979). Variants of polar ordination. *Vegetatio* 40:147—153.
- Gauch, H. G. and E. L. Stone (1979). Vegetation and soil pattern in a mesophytic forest at Ithaca, New York. *American Midland Naturalist* 102:332—345.
- Gauch, H. G. and T. R. Wentworth (1976). Canonical correlation analysis as an ordination technique. *Vegetatio* 33:17—22.
- Gauch, H. G. and R. H. Whittaker (1972). Comparison of ordination techniques. *Ecology* 53:868—875.
- Gauch, H. G., G. B. Chase, and R. H. Whittaker (1974). Ordination of vegetation samples by Gaussian species distributions. *Ecology* 55:1382—1390.
- Gauch, H. G., R. H. Whittaker, and T. R. Wentworth (1977). A comparative study of reciprocal averaging and other ordination techniques. *Journal of Ecology* 65:157—174.
- Gause, G. F. (1934). *Struggle for Existence*. Hafner, New York.
- George, D. G. and R. W. Edwards (1976). The effect of wind on the distribution of chlorophyll a and crustacean plankton in a shallow eutrophic reservoir. *Journal of Applied Ecology* 13:667—690.
- Gerdol, R., C. Ferrari, and F. Piccoli (1985). Correlation between soil characters and forest types: A study in multiple discriminant analysis. *Vegetatio* 60:49—56.
- Gibson, D. J. and P. Greig-Smith (1986). Community pattern analysis: a method for quantifying community mosaic structure. *Vegetatio* 66:41—47.
- Giller, P. S. (1984). *Community Structure and the Niche*. Chapman and Hall, London.
- Gittens, R. (1965). Multivariate approaches to a limestone grassland. I. A stand ordination. *Journal of Ecology* 53:385—401.
- Gittens, R. (1985). *Canonical Analysis: A Review with Applications in Ecology*. Springer—Verlag, Berlin.
- Gladfelter, W. B. and W. S. Johnson (1983). Feeding niche separation in a guild of tropical reef fishes (Holocentridae). *Ecology* 64:552—563.
- Gleason, H. A. (1926). The individualistic concept of the plant association. *Bulletin Torrey Botanical Club* 53:7—26.

- Goff, F. G. (1975). Comparison of species ordinations resulting from alternative indices of interspecific association and different numbers of included species. *Vegetatio* 31; 1—14.
- Goodall, D. W. (1953). Objective methods for the classification of vegetation. I. The use of positive interspecific correlation. *Australian Journal of Botany* 1; 39—63.
- Goodall, D. W. (1954a). Minimal area: a new approach. VIIth International Botanical Congress. Rapp. Comm. Parv. avant le Congress, Section 7, Ecologie, pp. 19—21.
- Goodall, D. W. (1954b). Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. *Australian Journal of Botany* 2; 304—324.
- Goodall, D. W. (1963). Pattern analysis and minimal area—some further comments. *Journal of Ecology* 51; 705—710.
- Goodall, D. W. (1964). A probabilistic index. *Nature*, London 203; 1098.
- Goodall, D. W. (1966). A new similarity index based on probability. *Biometrics* 22; 882—907.
- Goodall, D. W. (1970). Statistical plant ecology. *Annual Review of Ecology and Systematics* 1; 99—124.
- Goodall, D. W. (1973). Sample similarity and species correlation. In *Ordination and Classification of Communities* (R. H. Whittaker, Ed). W. Junk, The Hague, pp. 105—156.
- Goodall, D. W. (1974). A new method for the analysis of spatial pattern by random pairing of quadrats. *Vegetatio* 29; 135—146.
- Goodall, D. W. (1978a). Numerical classification. In *Classification of Plant Communities* (R. H. Whittaker, Ed). W. Junk, The Hague, pp. 247—286.
- Goodall, D. W. (1978b). Sample similarity and species correlation. In *Ordination of Plant Communities* (R. H. Whittaker, Ed.). W. Junk, The Hague, pp. 99—149.
- Goodall, D. W. and N. E. West (1979). A comparison of techniques for assessing dispersion patterns. *Vegetatio* 40; 15—28.
- Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53; 325—338.
- Gray, J. S. (1981). Detecting pollution induced changes in communities using the log—normal distribution of individuals among species. *Marine Pollution Bulletin* 12; 173—176.
- Green, R. H. (1966). Measurement of non—randomness in spatial distributions. *Researches in Population Ecology* 8; 1—7.
- Green, R. H. (1979). *Sampling Design and Statistical Methods for Environmental Biologists*. Wiley, New York.
- Green, R. H. (1980). Multivariate approaches in ecology: the assessment of ecological similarity. *Annual Review of Ecology and Systematics* 11; 1—14.
- Greig—Smith, P. (1952a). The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany* 16; 293—316.
- Greig—Smith, P. (1952b). Ecological observations on degraded and secondary forests in Trinidad, British West Indies. *Journal of Ecology* 40; 316—330.

- Greig-Smith, P. (1971). Analysis of Vegetation data: the user view—point. In *Statistical Ecology* (G. P. Patil, E. C. Pielou, and W. E. Waters, Eds.), Pennsylvania State University Press, University Park, PA, Volume 3, pp. 149—166.
- Greig-Smith, P. (1979). Presidential address 1979: Pattern in vegetation. *Journal of Ecology* 67: 755—780.
- Greig-Smith, P. (1983). *Quantitative Plant Ecology*, 3rd ed. University of California Press, Berkeley, CA.
- Gulmon, S. L. and H. A. Mooney (1977). Spatial and temporal relationships between two desert shrubs, *Atriplex hymenolytra* and *Tidestromia oblongifolia* in Death Valley, California. *Journal of Ecology* 65: 831—838.
- Hajdu, L. J. (1981). Graphical comparison of resemblance measures in phytosociology. *Vegetatio* 48: 47—59.
- Hartigan, J. A. (1975). *Clustering Algorithms*. Wiley, New York.
- Heip, C. (1974). A new index measuring evenness. *Journal of Marine Biological Association* 54: 555—557.
- Heltshe, J. F. and J. DiCanzio (1985). Power study of jackknifed diversity indices to detect change. *Journal of Environmental Management* 21: 331—341.
- Heltshe, J. F. and N. E. Forrester (1983a). Estimating species richness using the jackknife procedure. *Biometrics* 39: 1—11.
- Heltshe, J. F. and N. E. Forrester (1983b). Estimating diversity using quadrat sampling. *Biometrics* 39: 1073—1076.
- Heltshe, J. F. and N. E. Forrester (1985). Statistical evaluation of the jackknife estimate of diversity when using quadrat samples. *Ecology* 66: 107—111.
- Hendrickson, J. A. (1979). The biological motivation for abundance models. In *Statistical Distributions in Ecological Work* (J. K. Ord, G. P. Patil, and C. Taillie, Eds.), International Co-operative Publ., Fairland, MD, pp. 263—274.
- Hill, M. O. (1973a). The intensity of spatial pattern in plant communities. *Journal of Ecology* 61: 225—236.
- Hill, M. O. (1973b). Diversity and evenness: A unifying notation and its consequences. *Ecology* 54: 427—432.
- Hill, M. O. (1973c). Reciprocal averaging: An eigenvector method of ordination. *Journal of Ecology* 61: 237—249.
- Hill, M. O. (1974). Correspondence analysis: A neglected multivariate method. *Journal of the Royal Statistical Society. Series C* 23: 340—354.
- Hill, M. O. and H. G. Gauch (1980). Detrended correspondence analysis, an improved ordination technique. *Vegetatio* 42: 47—58.
- Hill, R. S. (1980). A stopping rule for partitioning dendrograms. *Botanical Gazette* 141: 321—324.

- Holgate, P. (1965). Some new tests of randomness. *Journal of Ecology* 53:261—266.
- Horn, H. (1966). Measurement of "overlap" in comparative ecological studies. *American Naturalist* 100:419—424.
- Hubalek, Z. (1982). Coefficients of association and similarity based on binary (presence—absence) data; an evaluation. *Biological Reviews* 57:669—689.
- Hubbell, S. P. (1979). Tree dispersion, abundance, and diversity in a tropical dry forest. *Science* 203:1299—1309.
- Huhta, V (1979). Evaluation of different similarity indices as measures of succession in arthropod communities of the forest floor after clear—cutting. *Oecologia* 41:11—23.
- Hurlbert, S. H. (1969). A coefficient of interspecific association. *Ecology* 50:1—9.
- Hurlbert, S. H. (1971). The non—concept of species diversity: A critique and alternative parameters. *Ecology* 52:577—586.
- Hurlbert, S. H. (1978). The measurement of niche overlap and some relatives. *Ecology* 59:67—77.
- Hurlbert, S. H. (1982). Notes on the measurement of overlap. *Ecology* 63:252—253.
- Hurlbert, S. H. and J. O. Keith (1979). Distribution and spatial patterning of flaminggos in the Andean Altiplano. *Auk* 96:328—342.
- Hutchinson, G. E. (1953). The concept of pattern in ecology. *Proceedings Academy Natural Sciences, Philadelphia, PA.*
- Hutchinson, G. E. (1959). Homage to Santa Rosalia, or why are there so many kinds of animals? *American Naturalist* 93:145—159.
- Ihm, P. and H. van Groenewoud (1975). A multivariate ordering of vegetation data based on Gaussian type gradient response curves. *Journal of Ecology* 63:767—777.
- Jaccard, P. (1908). Nouvelles recherches sur la distribution florale. *Bulletin Society Science Naturelle* 44:223—270.
- James, F. C. and S. Rathbun (1981). Rarefaction, relative abundance, and diversity of avian communities. *Auk* 98:785—800.
- Janson, S. and J. Vegelius (1981). Measures of ecological association. *Oecologia* 49:371—376.
- Jeglum, J. K., C. F. Wehrhahn, and J. M. A. Swan (1971). Comparisons of environmental ordinations with principal components—Vegetational ordinations for sets of data having different complexities. *Canadian Journal of Forest Research* 1:99—112.
- Jensen, D. D., G. B. Beus, and G. Storm (1968). Simultaneous statistical tests on categorical data. *Journal of Experimental Education* 36:46—56.
- Jensen, S. (1978). Influences of transformation of cover values on classification and ordination of lake vegetation. *Vegetatio* 37:19—31.
- Jensen, S. (1979). Classification of lakes in southern Sweden on the basis of their macrophyte composition by means of multivariate methods. *Vegetatio* 39:129—146.
- Jesberger, J. A. and J. W. Sheard (1973). A quantitative study and multivariate analysis of

- corticolous lichen communities in the southern boreal forest of Saskatchewan. *Canadian Journal of Botany* 51:185—201.
- Johnson, R. A. and D. W. Wichern (1982). *Applied Multivariate Statistical Analysis*. Prentice—Hall, Englewood Cliff, NJ.
- Johnson, R. B. and W. J. Zimmer (1985). A more powerful test for dispersion using distance measurements. *Ecology* 66:1084—1085.
- Johnson, R. W. and D. W. Goodall (1979). A maximum likelihood approach to non—linear ordination. *Vegetatio* 41:133—142.
- Kaesler, R. L. and E. E. Herricks (1976). Analysis of data from biological surveys of streams; diversity and sample size. *Water Resources Bulletin* 12:125—135.
- Kempton, R. A. and L. R. Taylor (1978). The Q—statistic and the diversity of floras. *Nature* 275:252—253.
- Kenkel, N. C. and L. Orloci (1986). Applying metric and nonmetric multidimensional scaling to ecological studies; some new results. *Ecology* 67:918—928.
- Kershaw, K. A. (1961). Association and co—variance analysis of plant communities. *Journal of ecology* 49:643—655.
- Kershaw, K. A. (1973). *Quantitative and Dynamic Plant Ecology*, 2nd ed. Elsevier, New York.
- Kessell, S. R. and R. H. Whittaker (1976). Comparisons of three ordination techniques. *Vegetatio* 32:21—29.
- Kolasa, J. and E. Biesiadka (1984). Diversity concept in ecology. *Acta Biotheoretica* 33:145—162.
- Knapp, R. (Ed.) (1984). *Sampling Methods and Taxon Analysis in Vegetation Science*. W. Junk, The Hague.
- Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29:115—129.
- Lamacraft, R. R., M. H. Friedel, and V. H. Chewings (1983). Comparison of distance based density estimates for some arid rangeland vegetation. *Australian Journal of Ecology* 8:181—187.
- Lambert, J. M. and W. T. Williams (1962). Multivariate methods in plant ecology. IV. Nodal analysis. *Journal of Ecology* 50:775—802.
- Lamont, B. B. and J. E. D. Fox (1981). Spatial pattern of six sympatric leaf variants and two size classes of *Acacia aneura* in a semi—arid region of Western Australia. *Oikos* 37:73—79.
- Lamont, B. B., S. Downs, and J. E. D. Fox (1977). Importance—value curves and diversity indices applied to a species—rich heathland in Western Australia. *Nature* 265:438—441.
- Lance, G. N. and W. T. Williams (1967). A general theory for classificatory sorting strategies. 1. Hierarchical systems. *Computer Journal* 9:373—380.
- Lance, G. N. and W. T. Williams (1968). A general theory for classificatory sorting strategies. 2. Clustering systems. *Computer Journal* 10:271—276.
- Lawlor, L. P. (1980). Overlap, similarity and competition coefficients. *Ecology* 61:245—251.

- Lawson, G. W. (1978). The distribution of seaweed floras in the tropical and subtropical Atlantic Ocean: A quantitative approach. *Botanical Journal of the Linnean Society* 76: 177—197.
- Lehman, J. T. and D. Scavia (1982). Microscale patchiness of nutrients in plankton communities. *Science* 216: 729—730.
- Legendre, L. and P. Legendre (1983). *Numerical Ecology*. Elsevier, New York.
- Legendre, P., S. Dallot, and L. Legendre (1985). Succession of species within a community: Chronological clustering, with applications to marine and freshwater zooplankton. *American Naturalist* 125: 257—288.
- Levins, R. (1968) *Evolution in Changing Environments: Some Theoretical Explorations*. Princeton University Press, Princeton, NJ.
- Lewis, T. G. and J. W. Doerr (1976). *Minicomputers: Structure and Programming*. Hayden Book, Rochelle Park, NJ.
- Livingston, R. J. (1976). Diurnal and seasonal fluctuations of organisms in a north Florida estuary. *Estuarine and Coastal Marine Science* 4: 373—400.
- Livingston, R. J., G. J. Kobylinski, F. G. Lewis, and P. F. Sheridan (1976). Long-term fluctuations of epibenthic fish and invertebrate populations in Apalachicola Bay, Florida. *Fishery Bulletin* 74: 311—321.
- Lloyd, M. (1967). Mean crowding. *Journal of Animal Ecology* 36: 1—30.
- Ludwig, J. A. (1979). A test of different quadrat variance methods for the analysis of spatial pattern. In *Spatial and Temporal Analysis in Ecology* (R. M. Cormack and J. K. Ord, Eds.). International Co-operative Publ., Fairland, MD, pp. 289—304.
- Ludwig, J. A. and D. W. Goodall (1978). A comparison of paired—with blocked—quadrat variance methods for the analysis of spatial pattern. *Vegetatio* 38: 49—59.
- Lyons, N. I. (1981). Comparing diversity indices based on counts weighted by biomass or other importance values. *American Naturalist* 118: 438—442.
- Madgwick, H. A. I. and P. A. Desrochers (1972). Association—analysis and the classification of forest vegetation of the Jefferson National Forest. *Journal of Ecology* 60: 285—292.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Science of India* 12: 49—55.
- Majer, J. D. (1976). The influence of ants and ant manipulation on the cocoa farm fauna. *Journal of Applied Ecology* 13: 157—175.
- Margalef, R. (1958). Information theory in ecology. *General Systematics* 3: 36—71.
- Martinka, R. R. (1972). Structural characteristics of blue grouse territories in southeastern Montana. *Journal of Wildlife Management* 36: 498—510.
- Mason, C. F. (1977). The performance of a diversity indice in describing the zoobenthos of two lakes. *Journal of Applied Ecology* 14: 363—367.
- Matta, J. F. and H. G. Marshall (1984). A multivariate analysis of phytoplankton assemblages

- in the western North Atlantic. *Journal of Plankton Research* 6:663—675.
- Matthews, J. A. (1978). An application of non—metric multidimensional scaling to the construction of an improved species plexus. *Journal of Ecology* 66:157—173.
- May, R. M. (1975). Pattern of species abundance and diversity. In *Ecology and Evolution of Communities* (M. L. Cody and J. M. Diamond, Eds.). Belnap Press, Cambridge, MA, pp. 81—120.
- May, R. M. (1981). Patterns in multi—species communities. In *theoretical Ecology* (R. M. May, Ed.). Sinauer Assoc. , Sunderland, MA, pp. 197—227.
- McCulloch, C. E. (1985). Variance tests for species association. *Ecology* 66:1676—1681.
- McDonald, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika* 27:397—415.
- McDonald, R. P. (1967). Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika* 32:77—112.
- McGuinness, K. A. (1984). Equations and explanationa in the study of species—area curves. *Biological Reviews* 59:423—440.
- McIntosh, R. P. (1967). The continuum concept of vegetation. *Botanical Reviews* 33:130—187.
- McIntosh, R. P. (1978). Matrix and plexus techniques. In *Ordination of Plant Communities* (R. H. Whittaker, Ed.). W. Junk, The Hague, pp. 151—184.
- McNaughton, S. J. and L. L. Wolf (1970). Dominance and the niche in ecological systems. *Science* 167:131—139.
- Menhinick, E. F. (1964). A comparison of some species—individuals diversity indices applied to samples of field insects. *Ecology* 45:859—861.
- Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* 69:89—107.
- Minshall, G. W. , R. C. Petersen, Jr. , and C. F. Nimz (1985). Species richness in streams of different size from the same drainage basin. *American Naturalist*. 125:16—38.
- Monk, C. D. (1967). Tree species diversity in the eastern deciduous forest with particular reference to north central Florida. *American Naturalist* 101:173—187.
- Morisita, M. (1959). Measuring of interspecific association and similarity between communities. *Memoirs Faculty Kyushu University. Series E* 3:65—80.
- Morisita, M. (1971). Composition of the I_d index. *Researches in Population Ecology* 13:1—27.
- Mueller, L. D. and L. Altenberg (1985). Statistical inference on measures of niche overlap. *Ecology* 66:1204—1210.
- Mueller—Dombois, D. (1974). *Aims and Methods of Vegetation Ecology*. Wiley, New York.
- Myers, J. H. (1978). Selecting a measure of dispersion. *Environmental Entomology* 7:619—621.
- Myers, W. L. and R. L. Shelton (1980). *Survey Methods for Ecosystem Management*. Wiley, New York.
- Nash, C. B. (1950). Association between fish species in tributaries and shore waters of

- western Lake Erie. *Ecology* 31:561—566.
- Nelder, J. A. (1975). *Computers in Biology*. Wykeham Publishing, London.
- Newsome, R. D. and R. L. Dix (1968). The forests of the Cypress Hills, Alberta and Saskatchewan, Canada. *American Midland Naturalist* 80:118—185.
- Niering, W. A. and C. H. Lowe (1984). Vegetation of the Santa Catalina Mountains: community types and dynamics. *Vegetatio* 58:3—28.
- Norris, J. M. and J. P. Barkham (1970). A comparison of some Cotswold beechwoods using multiple—discriminant analysis. *Journal of Ecology* 58:603—619.
- Noy—Meir, I. (1970). Component analysis of semi—arid vegetation in southeastern Australia. Ph. D. Dissertation, Australian National University, Canberra.
- Noy—Meir, I. (1971). Multivariate analysis of the semi—arid vegetation in southeastern Australia: Nodal ordination by component analysis. *Proceedings of the Ecological Society of Australia*. 6:159—193.
- Noy—Meir, I. (1973). Data transformations in ecological ordinations. I. Some advantages of non—centering. *Journal of Ecology* 61:329—341.
- Noy—Meir, I. (1974). Catenation: Quantitative methods for the definition of coenoclines. *Vegetatio* 29:89—99.
- Noy—Meir, I. and M. P. Austin (1970). Principal component ordination and simulated vegetation data. *Ecology* 51:193—215.
- Noy—Meir, I. and R. H. Whittaker (1977). Continuous multivariate methods in community analysis: Some problems and developments. *Vegetatio* 33:79—98.
- Noy—Meir, I. and R. H. Whittaker (1978). Recent developments in continuous multivariate techniques. In *ordination of Plant Communities* (R. H. Whittaker, Ed.). W. Junk, The Hague, pp. 337—378.
- Noy—meir, I., D. Walker, and W. T. Williams (1975). Data transformations in ecological ordinations. II. On the meaning of data standardization. *Journal of Ecology* 63:779—800.
- Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found In Japan and its neighbouring regions. *Bulletin Japanese Soc. Sci. Fisheries* 22:526—530.
- Oksanen, J. (1983). Ordination of boreal heath—like vegetation with principal component analysis, correspondence analysis, and multidimensional scaling. *Vegetatio* 52:181—189.
- Orloci, L. (1966). Geometric models in ecology. I. the theory and application of some ordination methods. *Journal of Ecology* 54:193—215.
- Orloci, L. (1967a). An agglomerative method for classification of plant communities. *Journal of Ecology* 55:193—206.
- Orloci, L. (1967b). Data centering: A review and evaluation with reference to component analysis. *Systematic Zoology* 16:208—212.
- Orloci, L. (1972). On objective functions of phytosociological resemblance. *American Midland Naturalist* 88:28—55.

- Orloci, L. (1973). Ordination by resemblance matrices. In *Ordination and Classification* (R. H. Whittaker, Ed.). W. Junk, The Hague, pp. 251—286.
- Orloci, L. (1974a). On information flow in ordination. *Vegetatio* 29:11—16.
- Orloci, L. (1974b). Revisions for the Bray and Curtis ordination. *Canadian Journal of Botany* 52:1773—1776.
- Orloci, L. (1978). *Multivariate Analysis in Vegetation Research*, 2nd ed. W. Junk, The Hague.
- Orloci, L. (1980). An algorithm for predictive ordination. *Vegetatio* 42:23—26.
- Orloci, L. and N. C. Kenkel (1985). *Introduction to Data Analysis with Examples from Population and Community Ecology. Statistical Ecology Monographs, Volume 1*, International Co-operative Publishing House, Fairland, MD.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, Sixth Series 2:559—572.
- Peet, R. K. (1974). The measurement of species diversity. *Annual Review of Ecology and Systematics* 5:285—307.
- Peet, R. K. (1975). Relative diversity indices. *Ecology* 56:496—498.
- Peet, R. K. and O. L. Loucks (1977). A gradient analysis of southern Wisconsin forests. *Ecology* 58:485—499.
- Pemberton, S. G. and R. W. Frey (1984). Quantitative methods in ichnology: spatial distribution among populations. *Lethaia* 17:33—49.
- Petratis, P. S. (1979). Likelihood measures of niche breadth and overlap. *Ecology* 60:703—710.
- Petratis, P. S. (1985). The relationship between likelihood niche measures and replicated tests for goodness of fit. *Ecology* 66:1983—1985.
- Phillips, D. L. (1978). Polynomial ordination: Field and computer simulation testing of a new method. *Vegetatio* 37:129—140.
- Pianka, E. R. (1973). The structure of lizard communities. *Annual Review of Ecology and Systematics* 4:53—74.
- Pielou, E. C. (1959). The use of point —to— plant distances in the study of the pattern of plant populations. *Journal of Ecology* 47:607—615.
- Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology* 13:131—144.
- Pielou, E. C. (1972a). Niche width and overlap: a method for measuring them. *Ecology* 53: 687—692.
- Pielou, E. C. (1972b). 2nd contingency tables in ecology. *Journal of Theoretical Biology* 34: 337—352.
- Pielou, E. C. (1974). *Population and Community Ecology*. Gordon and Breach, New York.
- Pielou, E. C. (1975). *Ecological Diversity*. Wiley, New York.
- Pielou, E. C. (1977). *Mathematical Ecology*. Wiley, New York.
- Pielou, E. C. (1979). *Biogeography*. Wiley, New York.

- Pielou, E. C. (1984). *The Interpretation of Ecological Data*. Wiley, New York.
- Pimental, R. A. (1979). *Morphometrics*. Kendall/Hunt, Dubuque, IA.
- Pimm, S. L. and D. P. Bartell (1980). Statistical model for predicting range expansion of the red imported fire ant, *Solenopsis invicta*, in Texas. *Environmental Entomology* 9: 653—658.
- Poole, R. W. (1974). *An Introduction to Quantitative Ecology*. McGraw—Hill, New York.
- Popma, J., L. Mucina, O. van Tongeren, and E. van der Maarel (1983). On the determination of optimal levels in phytosociological classification. *Vegetatio* 52: 65—75.
- Prentice, I. C. (1977). Non—metric ordination methods in ecology. *Journal of Ecology* 65: 85—94.
- Preston, F. W. (1948). The commonness, and rarity, of species. *Ecology* 29: 254—283.
- Preston, F. W. (1962). The canonical distribution of commonness and rarity, Part I. *Ecology* 43: 185—215.
- Pyke, G. H. (1982). Local geographic distributions of bumblebees near Crested Butte, Colorado; Competition and community structure. *Ecology* 63: 555—573.
- Quinn, J. F. and A. E. Dunham (1983). On hypothesis testing in ecology and evolution. *American Naturalist* 122: 602—617.
- Ratcliff, R. D. (1982). A correction of Cole's C7 and Hurlbert's C8 coefficients of interspecific association. *Ecology* 63: 1605—1606.
- Ratcliff, R. D. and R. D. Pieper (1981). Deciding final clusters: an approach using intra— and intercluster distances. *Vegetatio* 48: 83—86.
- Ratcliff, R. D. and R. D. Pieper (1982). Approaches to plant community classification for range managers. *Journal of Range Management Monograph Series*, No. 1.
- Ray, A. A. (Ed.) (1982). *SAS User's Guide; Statistic*, 1982 Edition. SAS Institute, Inc. Cary, NC.
- Revelante, N. and M. Gilmartin (1980). Microplankton diversity indices as indicators of eutrophication in the Northern Adriatic Sea. *Hydrobiologia* 70: 277—286.
- Rice, J., R. D. Ohmart, and B. W. Anderson (1983). Habitat selection attributes of an avian community; A discriminant analysis investigation. *Ecological Monographs* 53: 263—290.
- Ricklefs, R. E. and M. Lau (1980). Bias and dispersion of overlap indices: Results of some Monte Carlo simulations. *Ecology* 61: 1019—1024.
- Roberts, D. W. (1986). Ordination on the basis of fuzzy set theory. *Vegetatio* 66: 123—131.
- Rohlf, F. J. (1974). Methods of comparing classifications. *Annual Review of Ecology and Systematics* 5: 101—113.
- Rohlf, F. J. (1982). Consensus indices for comparing classifications. *Mathematical Biological Sciences* 59: 131—144.
- Rohlf, F. J. and R. R. Sokal (1981). *Statistical Tables*. Freeman. San Francisco, CA.
- Rohlf, F. J., J. Kishpaugh, and D. Kirk (1971). NT—SYS Numerical Taxonomy System

- of Multivariate Statistical Programs. Technical Report, State University of New York at Stony Brook, Stony Brook, NY.
- Romesburg, H. C. (1984). Cluster Analysis for Researchers. Lifetime Learning Publications, Belmont, CA.
- Root, R. B. (1967). The niche exploitation pattern of the blue—gray gnatcatcher. *Ecological Monographs* 37:317—350.
- Routledge, R. D. (1979). Diversity indices: Which ones are admissible? *Journal of Theoretical Biology* 76:503—515.
- Sanders, H. L. (1968). Marine benthic diversity: a comparative study. *American Naturalist* 102:243—282.
- Schluter, D. (1984). A variance test for detecting species association, with some example applications. *Ecology* 65:998—1005.
- Schoener, T. W. (1974). Resource partitioning in ecological communities. *Science* 185:27—39.
- Scott, D. and W. D. Billings (1964). Effects of environmental factors on the standing crop and productivity of an alpine tundra. *Ecological Monographs* 34:243—270.
- Searle, S. R. (1966). Matrix Algebra for the Biological Sciences. Wiley, New York.
- Shannon, C. E. and W. Weaver (1949). The Mathematical Theory of Communication. University Illinois Press, Urbana, IL.
- Sheldon, A. L. (1969). Equitability indices: dependence on the species count. *Ecology* 50:466—467.
- Shepard, R. N. (1980). Multidimensional scaling, tree—fitting, and clustering. *Science* 210:390—398.
- Shepard, R. N. and J. D. Carroll (1966). Parametric representation of non—linear data structures. In *Multivariate Analysis* (P. R. Krishnaiah, Ed.). Academic Press, New York.
- Simpson, E. H. (1949). Measurement of diversity. *Nature* 163:688.
- Sinclair, D. F. (1985). On tests of spatial randomness using mean nearest neighbour distance. *Ecology* 66:1084—1085.
- Smith, B. E. and G. Cottam. (1967). Spatial relationships of mesic herbs in southern Wisconsin. *Ecology* 48:546—558.
- Smith, E. P. (1984). A note on the general likelihood measure of overlap. *Ecology* 65:323—324.
- Sneath, P. H. A. and R. R. Sokal (1973). Numerical Taxonomy. Freeman, San Francisco, CA.
- Snedecor, G. W. and W. G. Cochran (1973). Statistical Methods. 6th ed. Iowa State University Press, Ames, IA.
- Sokal, R. R. (1974). Classification: purposes, principles, progress, prospects. *Science* 185:1115—1123.
- Sokal, R. R. and F. J. Rohlf (1981). Biometry, 2nd ed. Freeman, San Francisco, CA.
- Solomon, D. (1979). On a paradigm for mathematical modeling. In *Contemporary Quantitative Ecology and Related Ecometrics* (G. P. Patil and M. L. Rosenzweig, Eds.). International

- Co-operative Publishing, Fairland, MD.
- Southwood, T. R. E. (1978). *Ecological Methods: With Particular Reference to the Study of Insect Populations*. Wiley, New York.
- Spömer, G. G. (1973). The concepts of "interaction" and "operational environment" in environmental analyses. *Ecology* 54:200—204.
- Squires, E. R. and J. E. Klosterman (1981). Spatial patterning and competition in an aspen—white pine successional system. *American Journal of Botany* 68:790—794.
- Steele, R. G. D. and J. H. Torrie (1960). *Principles and Procedures of Statistics*. McGraw—Hill, New York.
- Stephenson, W., W. T. Williams, and S. D. Cook (1972). Computer analyses of Peterson's original data on bottom communities. *Ecological Monographs* 42:387—415.
- Stephenson, W., W. T. Williams, and G. N. Lance (1970). The macrobenthos of Moreton Bay. *Ecological Monographs* 40:459—494.
- Sugihara, G. (1980). Minimal community structure: An explanation of species abundance patterns. *American Naturalist* 116:770—787.
- Swan, J. M. A. and R. L. Dix (1966). The phytosociological structure of upland forest at Candle Lake, Saskatchewan. *Journal of Ecology* 54:13—40.
- Swanston, D. N., W. R. Meehan, and J. A. McNutt (1977). A quantitative geomorphic approach to predicting productivity of pink and chum salmon streams in southeast Alaska. USDA Forest Service Research Paper PNW—227, Portland, OR.
- Ter Braak, C. J. F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167—1179.
- Ter Braak, C. J. F. (1987). The analysis of vegetation—environment relationships by canonical correspondence analysis. *Vegetatio* 69:69—77.
- Thompson, D. C. (1980). A classification of the vegetation of Boothia Peninsula and the Northern District of Keewatin, N. W. T. *Arctic* 33:73—99.
- Turkington, R. and P. B. Cavers (1979). Neighbour relationships in grass—legume communities. III. Development of pattern and association in artificial communities. *Canadian Journal of Botany* 57:2704—2710.
- Usher, M. B. (1975). Analysis of pattern in real and artificial plant populations. *Journal of Ecology* 63:569—586.
- Usher, M. B. (1983). Pattern in the simple moss—turf communities of the sub—Antarctic and maritime Antarctic. *Journal of Ecology* 71:945—958.
- Van Belle, G. and I. Ahmad (1974). Measuring affinity of distributions. In *Reliability and Biometrics* (F. Proschan and R. J. Serfling, Eds.). SIAM Publication. Philadelphia, PA, pp. 651—668.
- Vandermeer, J. (1981). *Elementary Mathematical Ecology*. Wiley, New York.
- Vilks, G., E. H. Anthony, and W. T. Williams (1970). Application of association—analysis to

- distribution studies of recent Foraminifera. *Canadian Journal Earth Sciences* 7:1462—1469.
- Walker, B. H. and C. F. Wehrhahn (1971). Relationships between derived vegetation gradients and measured environmental variables in Saskatchewan wetlands. *Ecology* 52:85—95.
- Walker, H. A., S. B. Saila, and E. L. Anderson (1979). Exploring data structure of New York Bight benthic data using post-collection stratification of samples, and linear discriminant analysis for species composition comparisons. *Estuarine and Coastal Science* 9:101—120.
- Waring, R. H. and J. Major (1964). Some vegetation of the California coastal redwood region in relation to gradients of moisture, nutrients, light and temperature. *Ecological Monographs* 34:167—215.
- Went, F. W. (1942). The dependence of certain annual plants on shrubs in southern California deserts. *Bulletin Torrey Botanical Club* 69:100—114.
- Whittaker, R. H. (1952). A study of summer foliage insect communities in the Great Smoky Mountains. *Ecological Monographs* 22:1—44.
- Whittaker, R. H. (1965). Dominance and diversity in land plant communities. *Science* 147:250—260.
- Whittaker, R. H. (1967). Gradient analysis of vegetation. *Biological Reviews* 42:207—264.
- Whittaker, R. H. (1970). The population structure of vegetation. In *Gesellschafts-morphologie* (R. Tuxen, Ed.). W. Junk, The Hague, pp. 39—59.
- Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon* 21:213—251.
- Whittaker, R. H. (Ed.) (1978a). *Ordination of Plant Communities*. W. Junk, The Hague.
- Whittaker, R. H. (Ed.) (1978b). *Classification of Plant Communities*. W. Junk, The Hague.
- Whittaker, R. H. and H. G. Gauch (1973). Evaluation of ordination techniques. In *Ordination and Classification of Vegetation* (R. H. Whittaker, Ed.). W. Junk, The Hague, pp. 288—321.
- Whittaker, R. H. and W. A. Niering (1964). Vegetation of the Santa Catalina Mountains, Arizona. *Academy of Science* 3:9—34.
- Whittaker, R. J. (1987). An application of detrended correspondence analysis and non-metric multidimensional scaling to the identification and analysis of environmental factor complexes and vegetation structures. *Journal of Ecology* 75:363—376.
- Wiegand, G. (1981). Application of multiple discriminant analysis on the analysis of the correlation between macrophyte vegetation and water quality in running waters of Central Europe. *Hydrobiologia* 79:91—100.
- Williams, B. K. (1983). Some observations on the use of discriminant analysis in ecology. *Ecology* 64:1283—1291.
- Williams, W. T. (Ed.) (1976). *Pattern Analysis in Agricultural Science*. Elsevier, New York.
- Williams, W. T. and M. B. Dale (1965). Fundamental problems in numerical taxonomy. *Advances in Botanical Research* 2:35—68.
- Williams, W. T. and J. M. Lambert (1959). Multivariate methods in plant ecology. I. Association analysis in plant communities. *Journal of Ecology* 47:83—101.

- Williams, W. T. and J. M. Lambert (1961). Multivariate methods in plant ecology. III. Inverse association analysis. *Journal of Ecology* 49:717—729.
- Wishart, D. (1969). FORTRAN II programs for 8 methods of cluster analysis (Clustan I). Kansas Geological Survey Computer Contribution, No. 38, Kansas City, KS.
- Wolda, H. (1981). Similarity indices, sample size, and diversity. *Oecologia* 50:296—302.
- Wolda, H., F. W. Fish, and M. Estribi (1983). Faunistics of Panamanian cockroaches (Blattaria). *Uttar Pradesh Journal of Zoology* 3:1—9.
- Yarrington, G. A. and W. G. E. Green (1966). The distributional pattern of crustose lichens on limestone cliffs at Rattlesnake point. Ontario. *The Bryologist* 69:450—461.
- Yapp, W. B. (1979). Specific diversity in woodland birds. *Field Studies* 5:45—58.
- Yule, G. U. (1912). On the methods of measuring association in attributes. *Journal of the Royal Statistical Society, London* 75:579—642.
- Zahl, S. (1977). Jackknifing an index of diversity. *Ecology* 58:907—913.
- Zaret, T. M. and A. S. Rand (1971). Competition in tropical stream fishes; support for the competitive exclusion principle. *Ecology* 52:336—342.
- Zaret, T. M. and E. P. Smith (1984). On measuring niches and not measuring them. In *Evolutionary Ecology of Neotropical Freshwater Fishes* (T. M. Zaret, Ed.). W. Junk, The Hague, pp. 127—137.

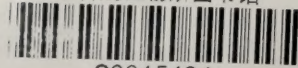


软盘程序 (BASIC) 说明:

本书软盘中的所有程序均是用 BASIC 写成的, 适用于 IBM-PCs 及兼容机, 以 ASCII 码方式存贮的, 使用前请留备份盘。

1. POISSON. BAS=与泊阿松分布一致的吻合性检验;
2. NEGBINOM. BAS=与负二项分布一致的吻合性检验;
3. BQV. BAS=区组样方方差法的空间格局分析;
4. PQV. BAS=所有可能成对的样方方差法空间格局分析;
5. TSQUARE. BAS=以 T-形取样距离为基础的空间格局分析;
6. LOGNORM. BAS=观测的多度数据与对数正态分布的配合;
7. RAREFRAC. BAS=种丰富度稀疏曲线方法;
8. SPDIVERS. BAS=丰富度、多样性和均匀指数;
9. SPOVRLAP. BAS=生态龕重叠的普遍和特定指数;
10. SPASSOC. BAS=种间关联的指数、卡方检验、方差比;
11. SPCOVAR. BAS=种亲和性的 R-方式相关系数;
12. SUDIST. BAS=取样单位相似性 Q-方式距离指数;
13. NASSOC. BAS=取样单位分类的关联分析;
14. CLUSTER. BAS=取样单位分类的聚类分析;
15. PO. BAS=极点排序;
16. PCA. BAS=主分量分析排序;
17. COA. BAS=对应分析排序;
18. DPC. BAS=非线性排序的无偏主分量分析;
19. NMDS. BAS=非线性排序的非度量多维调节;
20. SDA. BAS=分类结果简单判别分析;
21. PCREG. BAS=排序结果的一元和多元回归分析。

中科院植物所图书馆



S0015494

收到日期 98. 3. 20.

来源 赠书

书价 5.80

单据号

开票日期

统计生态学

——方法和计算入门

〔澳〕Jhon A. 拉德维格 著

〔美〕James F. 蓝诺兹 著

李育中 王炜 裴浩 译

刘钟龄 校

内蒙古大学出版社出版发行
中国农业科学院草原研究所印刷

开本: 787×1092/16 印张: 14.5 字数: 327 千

1991年4月第1版第1次印刷

印数: 1—1000 册

ISBN 7-81015-210-6/Q·10

定价: 5.80 元

26721

58.18
327

统计生态学——方法和计算入门

1990年

借者单位	借者姓名	借出日期	还书日期
------	------	------	------

	丁婉	28.4.20	
--	----	---------	--

58.18	29	2003年1月13日	
-------	----	------------	--

2003	2003年11月24日		
------	-------------	--	--

2003	2003年11月24日		
------	-------------	--	--

2003	2003年11月24日		
------	-------------	--	--

2003	2003年11月24日		
------	-------------	--	--

58.18
327

注 意

- 1 借书到期请即送还,
- 2 请勿在书上批改圈点,折角。
- 3 借去图书如有污损遗失等情形须照章赔偿。

26721

京卡 0701

责任编辑 呼 和

ISBN 7-81015-210-6/Q · 10

定 价： 5.80 元